

Ein- und mehrkanalige Ausfallsverschleierung für die digitale Drahtlosübertragung von Audiosignalen

Der Fakultät für Elektrotechnik und Informationstechnik
der Technischen Universität Graz

zur Erlangung des akademischen Grades

DOKTOR der technischen Wissenschaften

vorgelegt von

Dipl.-Ing. Cornelia Falch

Graz, 2007

Die Arbeit wurde am Institut für Elektronische Musik und Akustik der
Kunstuniversität Graz durchgeführt
und als Dissertation genehmigt von der Fakultät für Elektrotechnik und
Informationstechnik der Technischen Universität Graz.

Datum des Rigorosums: 28.02.2007
Vorsitzender: Univ.-Prof. Dipl.-Ing. Dr.techn. Anton Hofer
Begutachter: O.Univ.-Prof. Mag. Dipl.-Ing. Dr.techn. Robert Höldrich
Univ.-Prof. Dipl.-Ing. Dr.techn. Gernot Kubin

Inhalt

<i>Kurzfassung</i>	<i>iii</i>
<i>Abstract</i>	<i>iv</i>
KAPITEL 1 EINLEITUNG	5
1.1. MOTIVATION	5
1.2. ZIELSETZUNG	7
1.3. VERWANDTE ANWENDUNGSGEBIETE	11
1.4. GLIEDERUNG DER ARBEIT	14
KAPITEL 2 EINKANALIGE SYSTEME	17
2.1. FUNKTIONSWEISE DER AUSFALLSVERSCHLEIERUNG	18
2.2. SIGNALTHEORETISCHE GRUNDLAGEN	22
2.2.1. <i>Deterministische und stochastische Signale</i>	22
2.2.2. <i>Beschreibung eines Zufallsprozesses</i>	24
2.3. PRÄDIKTION	27
2.3.1. <i>Parametrische Modelle für nichtdeterministische Signale</i>	27
2.3.2. <i>Beziehung zwischen AR-Modellparameter und AKF</i>	29
2.3.3. <i>Lineare Prädiktion (LP)</i>	31
2.4. SUBSTITUTION UND INTERPOLATION	39
2.4.1. <i>Periodizitätsschätzung</i>	41
2.4.2. <i>Substitution</i>	46
2.4.3. <i>Interpolation mittels AR-Modell</i>	48
2.4.4. <i>Adaption der statistischen Interpolation in E-DROCON</i>	59
2.4.5. <i>Pufferverbesserung</i>	61
2.5. ZUSAMMENFASSUNG	62
KAPITEL 3 PERZEPTIVE EVALUIERUNG DER EINKANALIGEN VERSCHLEIERUNG	65

3.1.	TESTVORBEREITUNG	66
3.1.1.	<i>Auswahl der Versuchspersonen</i>	66
3.1.2.	<i>Testmethode</i>	67
3.1.3.	<i>Testmaterial, Audiosignale</i>	69
3.2.	DURCHFÜHRUNG DER TESTREIHEN	70
3.3.	SPEZIFIKATIONEN DER EINZELNEN TESTREIHEN	71
3.3.1.	<i>Testreihe 1: Extrapolation vs. Interpolation</i>	71
3.3.2.	<i>Testreihe 2: Parametersetting Extrapolation</i>	74
3.3.3.	<i>Testreihe 3: Periodizitätsschätzung</i>	75
3.4.	STATISTISCHE ANALYSE	75
3.4.1.	<i>Testreihe 1: Extrapolation vs. Interpolation</i>	77
3.4.2.	<i>Testreihe 2: Parametersetting Extrapolation</i>	83
3.4.3.	<i>Testreihe 3: Periodizitätsschätzung</i>	84
3.5.	ZUSAMMENFASSUNG	85
KAPITEL 4 MEHRKANALIGE SYSTEME		87
4.1.	STAND DER FORSCHUNG	89
4.2.	SPEKTRALES VERSCHLEIERUNGSFILTER	95
4.2.1.	<i>Basisalgorithmus</i>	97
4.2.2.	<i>Zeitglied zwischen Ziel- und Substitutionssignal</i>	103
4.2.3.	<i>Auswahl des Substitutionssignals</i>	106
4.3.	ZUSAMMENFASSUNG	110
KAPITEL 5 RESÜMEE UND AUSBLICK		113
	<i>Einkanalige Systeme</i>	114
	<i>Mehrkanalige Systeme</i>	116
ANHANG A PERZEPTIVE EVALUIERUNG		119
A.1.	BEWERTUNGSSKALA	119
A.2.	TESTREIHE 1	120
A.3.	TESTREIHE 2	124
A.4.	TESTREIHE 3	127
ANHANG B PATENTSCHRIFT P43136		131
ANHANG C ABKÜRZUNGSVERZEICHNIS		167
ANHANG D LITERATURVERZEICHNIS		169

Kurzfassung

Drahtlose Übertragungssysteme dienen als Basis für viele Produkte in der Elektroakustik und Audiotechnik. Dabei ist vor allem im professionellen Anwendungsbereich der Trend zu vollständig digitalisierten Ausführungen zu erkennen. Die digitale Drahtlosübertragung ist nicht resistent gegenüber Umwelteinflüssen, die entlang des Ausbreitungsmediums "Luft" auftreten können. Störungen können zum Totalausfall des Signals führen und müssen infolgedessen durch geeignete Techniken kompensiert werden. In realen echtzeitfähigen Systemen stellen die Ausfalllänge sowie der Zeitpunkt, an dem eine Unterbrechung auftritt, zwei unbekannte variable Parameter dar. Zur Rekonstruktion fehlender Datenblöcke werden Methoden untersucht, die einerseits auf die Eigenschaften der zu behandelnden Eingangssignale abgestimmt sind. Andererseits muss auch die Charakteristik der Fehler berücksichtigt werden. Als grundlegende Techniken eignen sich daher lineare, adaptive, extrapolierende Verfahren. Mithilfe eines adäquaten Modells ist es möglich, aus bekannten Signalanteilen eine Schätzung der ausgefallenen zu gewinnen. In dieser Arbeit werden verschiedene Modellrealisierungen wie beispielsweise das lineare Prädiktionsfehlerfilter für einkanalige, adaptive Filter für mehrkanalige sowie Substitutionstechniken für beide Arten der Ausfallsverschleierung präsentiert.

Abstract

Wireless transmission provides a basis for a number of products used in electro-acoustics and audio engineering. Particularly with professional equipment a trend towards complete digital systems is recognisable. Yet these wireless transmission systems are not resistant against interferences that occur along the radio link. Interferences may result in a total signal dropout and thus, must be compensated for incorporating appropriate concealment techniques. In real time systems, the size of a dropout as well as the beginning of a missing signal portion denote unknown variable parameters. The reconstruction of erroneous data is based on methods that on one hand take into consideration the signal's properties and of the other hand the characteristics of the dropouts. Proper fundamental techniques are linear, adaptive and extrapolating procedures. They apply an appropriately defined model to the known signal samples to generate an estimation of the unknown samples. This work investigates realisations like the linear prediction error filter for single-channel, adaptive filters for multi-channel and substitution techniques for both single- and multi-channel dropout concealment concepts.

Kapitel 1

Einleitung

1.1. Motivation

Die drahtlose Übertragung von Audiosignalen stellt seit der Markteinführung der Drahtlosmikrofone zu Beginn der 1990er Jahre ein wichtiges Forschungsgebiet in der Signalverarbeitung dar. Zur Zeit werden diese Produkte als Standardgeräte im Bereich von Bühnenaufführungen, Konzerten und Live-Shows eingesetzt. Während in den vergangenen Jahren die analogen Übertragungsformate im Vordergrund standen, stützt sich die Entwicklung der nächsten Jahre auf die Einführung von robusten und qualitativ hochwertigen digitalen Drahtlossystemen. Die Tendenz in die digitale Richtung hat mehrere Gründe. Zum einen stößt man in der analogen Technologie bereits an gewisse Grenzen, beispielsweise ermöglicht der digitale Bereich neben dem Transport und der Verarbeitung von Audiodaten zusätzlich die Verwertung von Metadaten. Diese können unter anderem Informationen über das Gesamtkonzept einer Bühneninstallation beinhalten. Die Idee, in zukünftigen Systemen die einzelnen Übertragungskanäle miteinander zu verbinden bzw. deren Interoperabilität auszunutzen, lässt sich ebenfalls durch den Einsatz digitaler Technologien realisieren. Parallel dazu unterstützt die rasante Entwicklung der zugrunde

liegenden Hardware in Bezug auf Rechenleistung und Speicherkapazität den Fortschritt von Softwarerealisierungen.

Im Allgemeinen ist die Methode der drahtlosen Signalübertragung nicht resistent gegenüber äußeren Beeinflussungen, die entlang der Übertragungsstrecke auftreten können. Bei digitalen Funkstrecken können Störungen unmittelbar zum Datenverlust und in weiterer Folge zum vollständigen Signalausfall führen. Die Verschlechterung der Signalqualität, akustisch wahrnehmbar als Knacksen oder Klicks, ist in jedem Fall inakzeptabel und durch geeignete Techniken zu kompensieren. Diese greifen auf Empfängerseite in den Signalweg ein und zielen darauf ab, die Ausfälle zu verschleiern. Neben der Gewährleistung der Qualität der Audiosignale entlang der Übertragung muss ein Verschleierungsmodul weitere Kriterien erfüllen. Als aktives Element im Signalpfad darf es seinerseits keine negativen Auswirkungen auf das Signal haben. Ein unerwünschter Effekt wäre unter anderem die zeitliche Verzögerung des Signals aufgrund von nicht optimalen Signalverarbeitungstechniken.

Folgende praktische Anwendungen verdeutlichen die Restriktion: Die Aufnahme eines Konzertes mit mehreren Mikrofonen ist zur Übertragung im Radio gedacht – die Latenzzeit der einzelnen Mikrofone ist nur dann unkritisch, wenn alle Geräte exakt dieselbe Signalverzögerung produzieren. Sind diese unterschiedlich, kann es zu Kammfiltereffekten oder zu Echos kommen. Die Aufnahme wird in jedem Fall unbrauchbar. Ähnliches gilt bei Live-Konzerten bzw. Fernsehübertragungen. Da die menschliche Wahrnehmung sehr empfindlich auf kombinierte audiovisuelle Information reagiert, ist eine exakte Synchronisation von Bild und Ton unabdingbar. Somit wird eine maximale Latenzzeit des gesamten Übertragungssystems von einigen Millisekunden gefordert.

1.2. Zielsetzung

Die vorliegende Arbeit beschäftigt sich mit der Entwicklung einer in das Gesamtsystem "digitale Drahtlosübertragung" integrierten Signalverarbeitungseinheit, die für die Zuverlässigkeit des Datentransportes verantwortlich ist. Dabei dient als Basis eine modifizierte Version des in [Horak 1998] beschriebenen Systems. Der vollständige Aufbau der zugrunde liegenden Übertragungskette inklusive der Ausfallsverschleierungseinheit ist in Abb. 1-1 dargestellt und umfasst für einen Kanal folgende Stufen: Sensor zur Signalaufnahme (Mikrofon), Analog-Digital-Konverter (ADC), Signalkompression und -kodierung, Sender, Übertragungskanal, Empfänger, Dekompression, Ausfallsverschleierung "DROCON". Am Ausgang dieses Moduls steht das Audiosignal in digitaler Form zur Verfügung – hier können weitere Funktionseinheiten angeschlossen werden, z. B. (Vor-)verstärker, Equalizer und dergleichen. Letztlich wird das Signal entweder auf einem digitalen oder analogen Speichermedium gespeichert oder es wird direkt über einen Lautsprecher wiedergegeben. In einer mehrkanaligen Anordnung, schematisch dargestellt in Abb. 1-2, wird das System um die entsprechende Anzahl an parallelen Kanälen erweitert, die anschließende Signalverwertung kann optional je Kanal oder in einer gemeinsamen Stufe ausgeführt werden.

Die Abkürzung DROCON steht für *dropout concealment* (Ausfallsverschleierung). Als eingebettetes System (*embedded system*) bildet DROCON ein aktives Element im Signalweg. Um ein möglichst hohes Maß an Flexibilität bei der Verwendung von DROCON zu erreichen, wurde es komplett unabhängig von der Sender/Empfänger-Einheit entwickelt und ist somit als eigenständiges Modul auch in andere Systeme integrierbar. Die Schnittstelle am Eingang von DROCON ist so definiert, dass das Übertragungsformat der digitalen Funkstrecke übernommen und zusätzlich ein Steuersignal in Form eines Statusbits mit der Information über den Zustand des Audiosignals (Ausfall ja/nein) vom Empfänger an DROCON gesendet wird. Das bedeutet, dass in der Ausfallsverschleierungseinheit selbst

keine spezielle Signalanalyse berücksichtigt werden muss, die zur Detektion fehlerhaft übertragener Datenpakete dient. Da die einzige Forderung am Ausgang von DROCON die Entkopplung zur nächsten Stufe ist, wird das fehlerfreie Signal im gleichen Audioformat zur Verfügung gestellt.

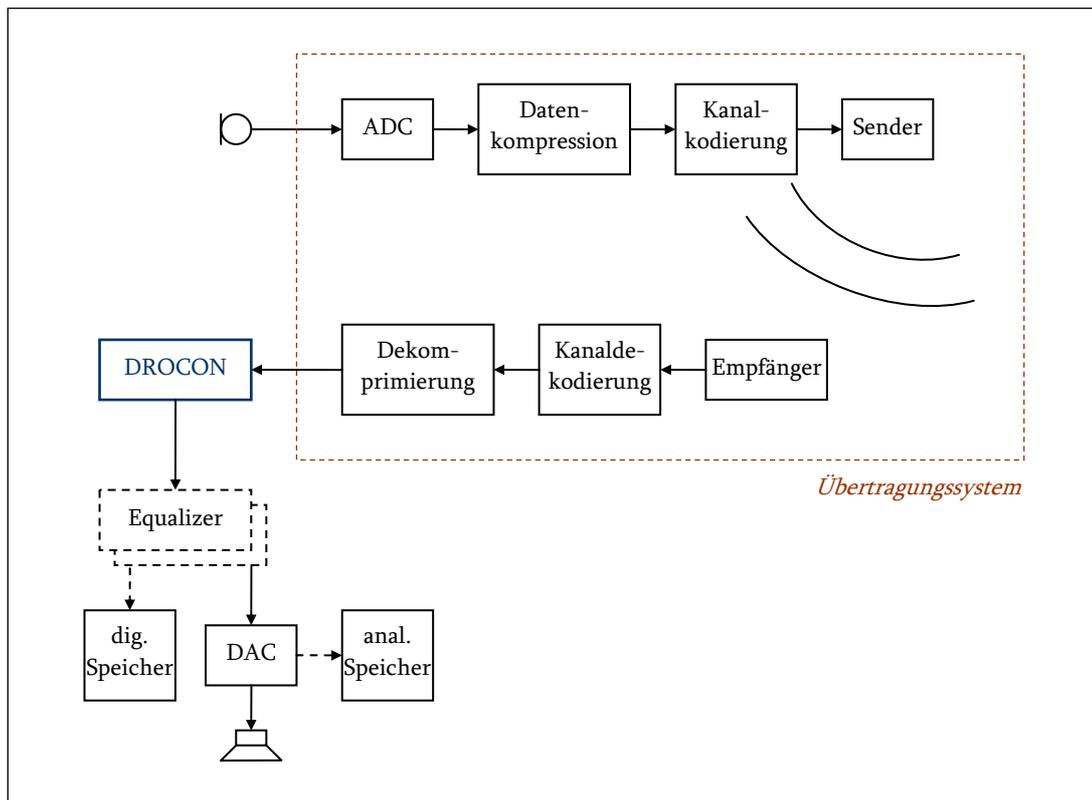


Abb. 1-1: Schema der Übertragungskette

Im Folgenden werden die wesentlichen Spezifikationen der Ausfallsverschleierung zusammengefasst:

- (1) Als Ausgangsmaterial dienen digitalisierte breitbandige Audiosignale wie Musik oder Sprache mit einer Abtastfrequenz von 44.1kHz bzw. 48kHz, deren Frequenzumfang damit den gesamten Hörbereich von 20Hz bis 20kHz einschließt. Die vorgeschaltete Übertragung erfolgt paketiert mit einer Paketgröße von wenigen Abtastpunkten, im konkreten Fall umfasst sie 4 Abtastpunkte.

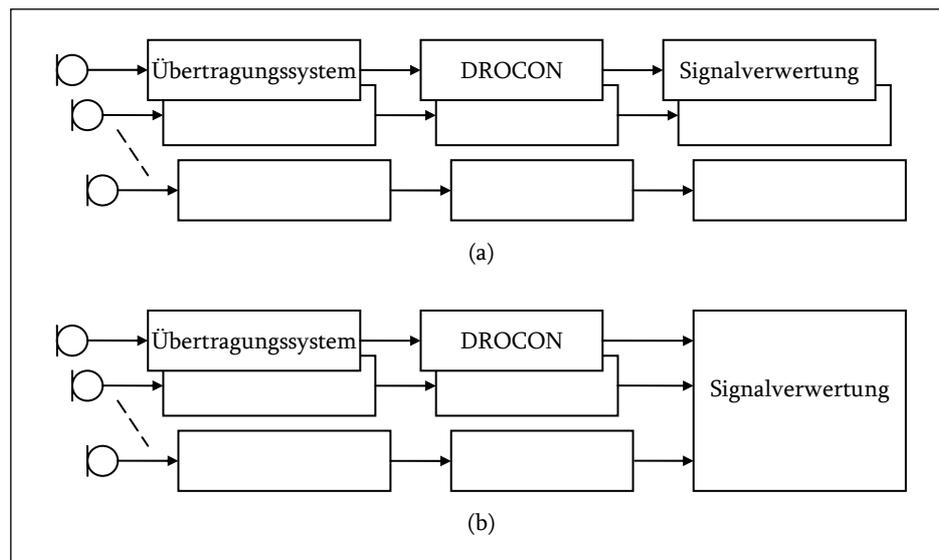


Abb. 1-2: Schema einer mehrkanaligen Übertragung
 (a) mit getrennter Verwertung der Einzelsignale,
 (b) mit gemeinsamer Verwertung aller Kanäle

- (2) Störungen entlang einer analogen Funkstrecke bewirken ein kurzzeitiges Ansteigen von Rauschen, das dem Audiosignal als Störung überlagert ist. Im Gegensatz dazu können sie bei der digitalen Drahtlosübertragung zum totalen Signalausfall führen. Die Charakteristik der Signalausfälle variiert zum Teil erheblich mit den verwendeten Methoden der Übertragung. Als Kenngrößen werden die maximale Länge der Ausfälle sowie deren statistische Häufigkeitsverteilung angegeben. Für die Entwicklung der DROCON-Algorithmen standen keine Vorstudien über das Fehlverhalten digitaler Drahtlosmikrofonsysteme zur Verfügung. Zu Testzwecken wurden folgende Grenzen festgelegt: die Ausfälle liegen zwischen einer und zehn Millisekunden, der Bereich von zwei bis sechs Millisekunden wurde insbesondere untersucht. Die Häufigkeit der Ausfälle wurde nicht genau spezifiziert, es sollten unterschiedliche Verteilungen evaluiert werden. Daten aus vergleichbaren Anwendungsgebieten wie beispielsweise GSM, WLAN etc. wurden nicht berücksichtigt.
- (3) Eine wesentliche Herausforderung an das Gesamtsystem stellt dessen Einsatz in einer Echtzeitumgebung dar. Wie bereits erwähnt sind zu große Latenzzeiten für ein qualitativ hochwertiges Endprodukt unzulässig. In praktischen Anwendungen hat sich gezeigt, dass eine gesamte Verzögerung von 4ms zwischen Signalaufnahme und –wiedergabe noch tolerierbar ist, [Horak 1998]. Da diese Latenzzeit bereits dem zugrunde liegenden

Übertragungssystem inhärent ist, wird in der vorliegenden Arbeit das extreme Ziel der "Null-Latenz" für die Verschleierungseinheit angestrebt. Es soll ein Algorithmus entwickelt werden, der keine zusätzliche Zeitverzögerung des Audiosignals hervorruft. Konkret bedeutet das:

Das Datenpaket, das zu einem Zeitpunkt t der Verschleierungseinheit zur Verfügung steht, wird (bei Ausfall) behandelt und unmittelbar der nachfolgenden Anwendung übergeben. Eine Pufferung der Daten im aktiven Signalweg ist nicht möglich, somit ist die Verzögerung einerseits von der Rechenkomplexität des Verschleierungsalgorithmus und andererseits von der Leistung des Signalprozessors abhängig.

Für die Generierung des geschätzten Zielsignals wird hingegen ein Ausschnitt der Signalvergangenheit benötigt, der ohne negative Auswirkungen auf den Signalfluss in einem Puffer gespeichert werden kann.

- (4) Etwas weniger strikte Vorgaben werden bezüglich Speicher- und Rechenaufwand festgelegt. Da eine gezielte Optimierung dieser Größen nicht vorrangiger Gegenstand dieser Arbeit war, wurden sie zwar berücksichtigt, jedoch nicht so kritisch gehandhabt wie die Latenzzeit. Eine Subbandaufteilung am Eingang der Verschleierungseinheit wurde beispielsweise wegen des zusätzlichen Rechenaufwandes und der mit der Filterung einhergehenden zusätzlichen Verzögerung nicht in Betracht gezogen. Die Eigenständigkeit der vorgeschlagenen Methoden ermöglicht eine Implementierung in die einzelnen Subbänder des Dekodierers¹, allerdings ist dabei darauf, dass unerwünschte Quereffekte das Leistungsverhalten des Gesamtsystems nicht negativ beeinflussen.

Diese Abhandlung gibt einen Einblick in die Problematik der Ausfallsverschleierung bei digitalen Signalen und dient unter anderem als Grundlage für weitere Forschungsarbeiten. Die Optimierung der vorgeschlagenen Algorithmen in einer realen Umgebung war nicht vordergründiges Ziel der Arbeit. Die Tatsache, dass eine Optimierung individuell für unterschiedliche Gesamtsysteme und Einsatzgebiete durchgeführt werden muss, unterstreicht die Ausgliederung aus dem Aufgabenbereich der vorliegenden Arbeit.

¹ Dies ist z. B. bei subbandbasierten Kodierverfahren von Bedeutung. Der als Basis zur Verfügung stehende Kodieralgorithmus besitzt diese Eigenschaft, welche allgemein bei digitalen Übertragungssystemen von Audiosignalen häufig Anwendung findet.

1.3. Verwandte Anwendungsgebiete

Die Klassifizierung von Fehlerverschleierungstechniken kann nach unterschiedlichen Gesichtspunkten vorgenommen werden. Eine grundlegende Einteilung ergibt sich durch die chronologische Entwicklung digitaler Audiosysteme: mit der Einführung digitaler Speichermedien wurde die Restauration von analogem Datenmaterial zu einem wichtigen Forschungsgebiet. Das Hauptziel bestand/besteht in der Verbesserung der qualitativ schlechten Analogaufnahmen und der Wiederherstellung von verlorengegangenen Signalblöcken (die z. B. durch Beschädigungen analoger Speichermedien entstehen) unter Ausnutzung der intakten Signalanteile vor und nach den gestörten/fehlenden Daten. Dazu eignen sich Prädiktions-, Interpolations- und Extrapolationsalgorithmen, die auf der Ermittlung von stochastischen, deterministischen oder kombinierten Signalmodellen beruhen. Am häufigsten zum Einsatz kommen stochastische Modelle, da bereits Methoden mit geringer Komplexität ein ausreichend gutes Leistungsverhalten besitzen (z. B. zeitbereichsbasierte Verfahren). Einige wichtige Literaturquellen, die sich ausführlich mit diesem Thema beschäftigen sind u. a. [Rayner 1991], [Niedzwiecki 1993], [Etter 1996], [Godsill 1997], [Karjalainen 1997], [Godsill 1998], [Härmä 2001a], [Härmä 2001b], [Kauppinen 2001], [Esquef 2002], [Esquef 2003], [Roth 2004], [Esquef 2006]. Die Verwendung von Sinusmodellen mit und ohne stochastischem Restsignal wird u. a. in [Maher 1994], [Parikh 2000], [Lagrange 2005] gezeigt. Diese Methoden besitzen im Allgemeinen einen aufwändigen Analyseteil zur Detektion der deterministischen Komponenten, um eine qualitativ hochwertige Resynthese zu erzielen.

Durch die Entwicklung von Komprimierungsalgorithmen einerseits sowie die laufende Verbesserung der Netzwerktechnologie in Bezug auf steigende Übertragungskapazitäten andererseits hat sich die digitale Audioübertragung (*digital audio streaming*) rasch zu einem wesentlichen Teilbereich in Internetanwendungen gemacht, z. B. Übertragung von Musikdateien, Internetradio, Internettelefonie und -konferenz, etc. Einwandfreier

Datentransport wird in solchen Systemen allgemein nicht garantiert. Deshalb ist die Erkennung und Kompensation von Übertragungsfehlern notwendig, wobei unterschiedliche Fehlerverschleierungstechniken zum Einsatz kommen. Grundsätzlich ist der Datenverkehr in Netzwerken über die zwei Protokolle TCP/IP bzw. UDP/IP geregelt², die beide die Strategie eines *best-effort service* verfolgen und keine Qualitätsgarantie (*quality of service*) des Transfers übernehmen [Rahkila 1997], [Xu 2000]. Während die TCP/IP-Architektur eine zuverlässige Datenzustellung anstrebt (der Sender erwartet vom Empfänger eine Rückmeldung über den Erhalt der Pakete), liegt das Ziel der UDP/IP-Architektur in einem möglichst schnellen Versand, ohne dessen Zuverlässigkeit zu überprüfen. Audio- und Videosignale werden vorwiegend mittels UDP/IP übertragen, da die deutlich geringere Latenzzeit (im Vergleich zu TCP/IP) besonders bei echtzeitfähigen Datenstrom-Anwendungen (*real-time streaming*) im Vordergrund steht. Zur Behandlung von Übertragungsfehlern (z. B. fehlerhafte Pakete, nicht korrekt oder zu spät zugestellte Pakete, etc.) sind zusätzliche Verschleierungseinheiten notwendig.

Die Echtzeitfähigkeit beim einseitigen Versand von Audio- und Videodaten, wie z. B. Internetradio, IPTV, etc., ist nicht sehr kritisch, weshalb eine Pufferung von mehreren Sekunden bis zu einer Minute zu Beginn der Übertragung möglich ist, um ein entsprechend hohes Maß an Robustheit gegenüber Signalverzögerungen zu erzielen. Weerackody et. al. [Weerackody 2001] entwickeln eine Methode zur Verringerung der sogenannten Startverzögerung (*start-up delay*). Die Anforderungen für interaktives Audio- und Videostreaming (z. B. *voice-over-IP*, *videoconferencing*, etc.) hingegen sind restriktiver, es werden nur minimale Zeitverzögerungen toleriert. Von der ITU (*International Telecommunication Union*) wird etwa eine maximale End-zu-End-Verzögerung von 150ms vorgeschlagen [ITU-T 2000], [Lindblom 2002]. Allerdings übersteigt auch dieser Grenzwert die erlaubten Limits bei Mikrofonsystemen um das 30- bis 40fache.

In Anlehnung an [Perkins 1998] kann eine weitere Klassifizierung der Verschleierungstechniken danach erfolgen, an welchem "physikalischen Ort" diese stattfindet. Perkins et. al. unterteilen Datenstrom-Anwendungen in sendergesteuerte und empfängerbasierte Techniken. Verfahren der ersten Gruppe sind beispielsweise

² IP: internet protocol; TCP: transmission control protocol; UDP: user data protocol.

mehrfachbeschreibende Kodieralgorithmen (*multiple description coding*) [Areal 2000], [Barros 2002], Verschachtelung (*interleaving*) [Wah 1998], [Wah 2000b], [Sinha 2003], die vorwärtsgerichtete Fehlerkorrektur (*forward error correction*) oder die wiederholte Übertragung von Datenpaketen (*retransmission*) [Perkins 1998]. Kennzeichen für diese Verfahren sind unter anderem eine erhöhte Redundanz und Bandbreite sowie Verzögerung, damit verbunden aber auch eine bessere Qualität der Verschleierung. Auf Empfängerseite werden vorwiegend Substitutions- und Interpolationsverfahren implementiert, in denen zum Teil ähnliche Prädiktionsalgorithmen eingesetzt werden wie bei der Signalrestauration. Einfachere Verfahren wie Ausblenden, Wiederholung von unmittelbar vorangegangenen Paketen oder dem Einfügen von Rauschen finden aufgrund ihrer nicht zufriedenstellenden Qualität kaum mehr Anwendung, [Herre 1995], [Schildbach 2002]. Speziell bei der Übertragung von Audiosignalen im MPEG-Format werden vermehrt Methoden ausgenutzt, die zwar empfängerseitig aber noch vor der Dekodierung eingebaut sind. An dieser Stelle liegen die Audiodaten als spektrale Abbildung (Linien der modifizierten diskreten Kosinustransformation, MDCT) vor, weshalb die Verschleierung in einer transformierten Domäne bzw. in Subbändern stattfindet, [Lauber 2000], [Lauber 2001], [Schildbach 2002], [Ofir 2005].

Eine dritte Möglichkeit der Klassifizierung richtet sich nach der maximal zulässigen Signalverzögerung, die durch die zusätzliche Verarbeitungsstufe erhöht wird. Die folgende hierarchische Struktur veranschaulicht, wie kritisch die Latenzzeit der Einheit zur Signalwiederherstellung, Fehler- oder Ausfallsverschleierung in praktischen Anwendungsgebieten ist:

Auswirkung der Latenzzeit	Max. zulässige Verzögerung des Gesamtsystems	Anwendung
unkritisch	-	Signalrestauration
wenig kritisch	bis zu einer Minute	einseitiger Datentransfer in IP-Netzwerken
kritisch	< 150ms	interaktiver Datenaustausch in IP-Netzwerken
sehr kritisch	< 4ms	drahtlose Mikrofonübertragung

Die dieser Arbeit zugrunde liegende Unterscheidung zwischen ein- und mehrkanaligen Verschleierungsmethoden findet aktuell in der Literatur kaum Erwähnung. Einige Arbeiten beschäftigen sich zwar speziell mit mehrkanaligen Systemen, eine allgemeine Zusammenfassung sowie ein direkter Vergleich zwischen den verschiedenen Ansätzen wurde jedoch bisher noch nicht durchgeführt. Die einfachste Form der mehrkanaligen Verschleierung ist in der US-Patentschrift 0039464 [Virolainen 2004] festgehalten, welche die Konservierung der räumlichen Information bei der Übertragung von Stereosignalen zum Ziel hat. Der Interkanal-Zusammenhang wird im Zeitbereich analysiert, als geeignete Techniken hierfür werden die Berechnung der Phasendifferenz, der Kreuzkorrelation oder der Tiefpassfilterung mit anschließender Maximumdetektion genannt. Einen Überblick über weitere mehrkanalige Verfahren, die nicht auf das Stereoformat beschränkt sind, bietet Abschnitt 4.1.

1.4. Gliederung der Arbeit

In Kapitel 2 bilden einkanalige Systeme den Gegenstand der Untersuchungen. Zur Behandlung von Monosignalen dienen Zeitbereichsverfahren wie Extrapolations- und Substitutionsalgorithmen. Anhand von informellen Hörtests wurde die Leistungsfähigkeit beider Methoden in Bezug auf perzeptive Audioqualität und Rechenkomplexität beurteilt. Eine geeignete Kombination daraus ermöglicht die Ausnutzung der jeweiligen Vorteile und führt in Folge zu einem optimierten Ansatz. Dieser beachtet eine spezielle Art der Interpolation, die ohne Pufferung des aktiven Signalflusses realisiert werden kann. Die in der Zukunft erwarteten Datenpunkte werden dabei der Vergangenheit entnommen und als "virtuelles" (zukünftiges) Signal interpretiert. Die Bestimmung des passenden Ausschnitts basiert auf bekannten Methoden des Mustervergleichs (*pattern matching*).

Kapitel 3 beschäftigt sich mit der perzeptiven Evaluierung der Qualität der Verschleierungsalgorithmen, die in Form von drei subjektiven Hörtests durchgeführt wurde. Diese beruhen auf den Richtlinien der "International Telecommunication Union", Sektion "Radiocommunication" (ITU-R). Die zugrunde liegenden fehlerbehafteten und fehlerverschleierte Audiobeispiele wurden vorab generiert, wobei in den einzelnen Testreihen jeweils leicht modifizierte Annahmen bzgl. Ausfallsstatistik getroffen wurden. Die Ziele der Untersuchungen beinhalten den Vergleich der vorgeschlagenen E-DROCON-Techniken, die Optimierung der Parametereinstellungen des Extrapolationsalgorithmus sowie die Verifizierung der in Betracht gezogenen Methoden der Periodizitätsschätzung.

Mehrkanalige Systeme, wie z. B. die Mikrofonierung von Orchesterkonzerten (siehe Abb. 1-3), sind Ausgangspunkt der in Kapitel 4 entwickelten M-DROCON-Methoden. Kennzeichnend für diese Techniken ist die Strategie, einen ausgefallenen Kanal aus einem oder mehreren alternativen Kanälen zu rekonstruieren. Entsprechende Verfahren müssen in der Lage sein, die intakten parallelen Kanäle zur Ausfallsverschleierung gemeinsam zu nutzen. Adaptive Filter, die weitverbreitet Anwendung in der Verarbeitung von MIMO-Systemen finden, wurden deshalb in der vorliegenden Arbeit untersucht, um sie zu einer leistungsfähigen Verschleierungsmethode zu führen. Die zum Teil sehr großen Distanzen (> 5m) zwischen den Mikrofonen sowie zwischen Instrumenten und Mikrofonen verhindern in Verbindung mit der meist halligen Umgebung die für die Adaption notwendige stabile Korrelation zwischen den Kanälen. Aus diesem Grund wurde ein neuartiges spektrales Verschleierungsfilter entwickelt, das aus dem Verhältnis der Einhüllenden der spektralen Betragsverläufe von Ziel- und Substitutionssignal geschätzt wird und somit auf Korrelationsschätzungen weitgehend verzichtet. Das Substitut kann aus einer gewichteten Überlagerung von Parallelkanälen zusammengesetzt sein. Die Vernachlässigung der Phasenbeziehungen wird entweder durch ein geeignetes Überblenden zu Beginn und am Ende der Signalersetzung oder durch die Verarbeitung der Laufzeitdifferenzen zwischen den betrachteten Signalen in einer separaten Einheit kompensiert.

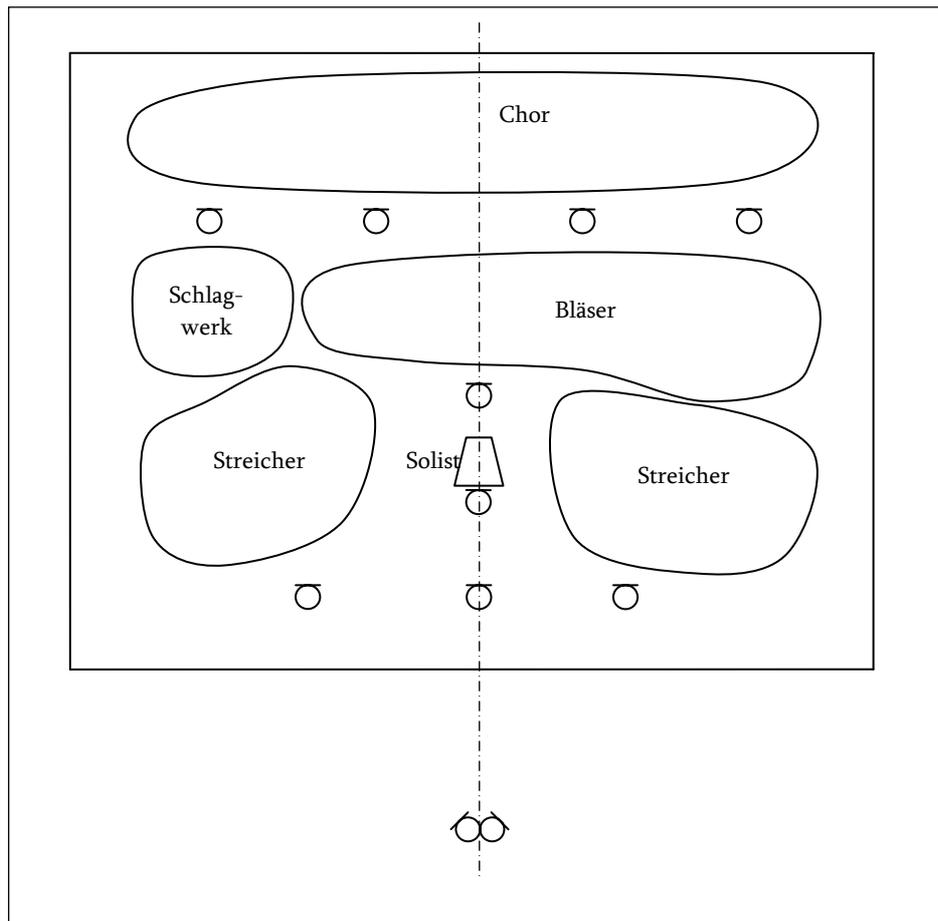


Abb. 1-3: Beispiel einer mehrkanaligen Mikrofonierung

Abschließend werden in Kapitel 5 die Ergebnisse der Arbeit zusammengefasst. Einige Möglichkeiten zur Fortsetzung und Vertiefung der gewonnenen Erkenntnisse werden im Ausblick diskutiert.

Kapitel 2

Einkanalige Systeme

Die Problematik der Fehler- oder Ausfallsverschleierung bei digitalen Audiosignalen wie Musik und/oder Sprache ist ein wichtiges Thema in der Signalverarbeitung. Aufgrund der allgemeinen Tendenz zur kompletten Digitalisierung von Audiosystemen entstehen ständig neue Anwendungsbereiche für Verschleierungstechniken. In der vorliegenden Arbeit wird die Strategie verfolgt, diese so in eine Echtzeitumgebung zu integrieren, dass keine Verzögerung im aktiven Signalweg induziert wird. Dadurch unterscheidet sie sich grundlegend von derzeit aktuellen Einsatzgebieten, in denen auch echtzeitfähige Systeme Latenzzeiten bis zu einigen hundert Millisekunden tolerieren.

Dieses Kapitel widmet sich einkanaligen Systemen, d. h. es werden im Folgenden ausschließlich Monosignale als gültige Eingangsfunktionen zugelassen. Zur Ausfallsverschleierung für diese Art von Signalen dienen zeitreichsbasierte Prädiktionsmethoden als Grundlage, wie sie beispielsweise in der Signalrestauration oder der Audioübertragung in Netzwerken auf Empfängerseite implementiert werden. In diesem Sinne wurde versucht, die bekannten Techniken weiterzuentwickeln und dahingehend zu modifizieren, dass sie das Kriterium der "Null-Latenz" erfüllen.

2.1. Funktionsweise der Ausfallsverschleierung

Vor der detaillierten Erläuterung von Theorie und spezieller Anwendung der mathematischen Methoden wird der allgemeine Aufbau des DROCON-Verfahrens definiert. In die in Abb. 2-1 dargestellte Rahmenstruktur sind die einzelnen Algorithmen eingebettet. Sie besteht aus vier Hauptzweigen und dem Block "*improve buffer*". Das Flussdiagramm ist auf die Implementierung des Gesamtsystems in die Programmiersprache MATLAB abgestimmt. In der realen Umgebung wird das Audiosignal paketierte mit einer Paketgröße von vier Abtastpunkten übertragen. Dieses Format wird im Programm beibehalten, die einzelnen Pakete werden mithilfe einer Schleife abgearbeitet.

Nach der Festlegung der Parameter und einer Initialisierungsphase wird in der ersten Entscheidung abgefragt, ob der aktuelle Block gültige Signaldaten enthält oder ob diese als ausgefallener Block gekennzeichnet ist. Die Bezeichnung "aktuell" bezeichnet jenen Block (mit dem Index m), der im Algorithmus tatsächlich, d. h. zum aktuellen Zeitpunkt, bearbeitet wird. Diese Klassifizierung ist von Bedeutung, da die MATLAB-Implementierung lediglich eine Simulation einer echtzeitfähigen Realisierung des Algorithmus darstellt. Sind die aktuellen Daten in Ordnung – *no DO in actual frame*³ – wird in der nächsten Entscheidung abgefragt, in welchem Zustand der Steuerparameter do im unmittelbar vorangegangenen Block war. Ist $do(m-1) = false$, so liegt kein Signalausfall vor – *no DO (frame correct)*, und das Eingangssignal wird unverändert an die Einheit zur Pufferverbesserung weitergeleitet. Im anderen Fall, $do(m-1) = true$, repräsentiert das aktuelle Datenfenster das erste gültige nach einem Ausfall – *first original frame after DO*. Die Bestimmung dieses Blocks ist notwendig, da ab diesem Zeitpunkt direkt mit dem Übergang vom geschätzten Zielsignal zum originalen begonnen wird. Im Anschluss erfolgt wiederum die Pufferverbesserung. Findet im aktuellen Block ein Ausfall statt – *DO in actual frame*, muss abermals der Zustand des letzten, also $do(m-1)$, überprüft werden. Der Beginn eines Ausfalls wird detektiert, wenn die aktuellen

³ Die Abkürzung DO steht für *dropout* (Ausfall).

Datensamples ungültig, aber die davor liegenden gültig sind. In diesem Fall muss die Ausfallsverschleierung das erste Mal einsetzen, d. h. im Block *first DO frame* wird diese gestartet.

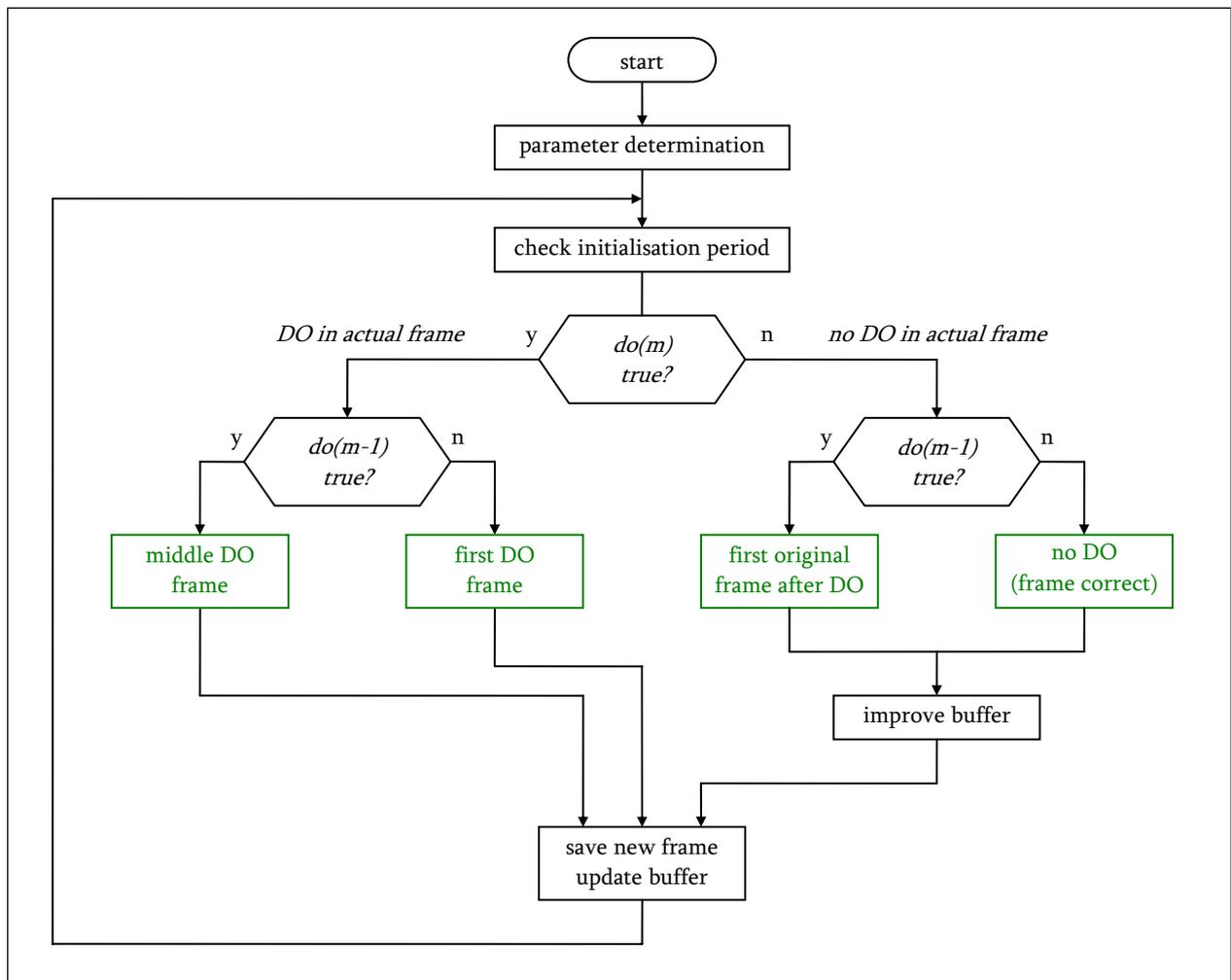


Abb. 2-1: Rahmenstruktur des DROCON Algorithmus

Als letzten möglichen Zustand kann sich das Signal gerade innerhalb eines Ausfalls befinden – *middle DO frame*. Das bedeutet, dass die Ausfallsverschleierung aktiv ist und das geschätzte Zielsignal zur Speicherung bzw. an den Ausgang der DROCON-Einheit weitergeleitet wird. Gleichzeitig wird das Signal auch zur Aktualisierung des Puffers im Block "improve buffer" verwendet. Ziel dieser Einheit ist die Verbesserung der im Puffer gespeicherten Signaldaten vor allem dann, wenn mehrere kurze Signalausfälle aufeinander folgen. Dies hat zur Folge, dass der Puffer eine hohe Variation zwischen Zielsignal und dessen Schätzung aufweist, was sich wiederum nachteilig auf nachfolgende Berechnungen der Ausfallsverschleierung

auswirkt. Die Verbesserung erfolgt im Nachhinein als Rückwärtsadaption mit einer klassischen Interpolationstechnik.

Der Puffer dient zur Speicherung eines bestimmten Ausschnitts der Signalvergangenheit. In der Echtzeitanwendung sind dem reinen Übertragungssystem Vergangenheit und Zukunft unbekannt und nur die jeweils aktuellen Daten verfügbar (im vorliegenden Fall entspricht dies exakt einem Block oder vier Samples). Da zur Überbrückung des Ausfalls jedoch ein Teil der Signalhistorie verwendet wird, wird diese in einem Puffer abgelegt, dessen Länge eingestellt werden kann. Wenn die Übertragung korrekt verläuft und das Signal nicht mit Ausfällen behaftet ist, wird der Puffer mit dem Zielsignal gefüllt.

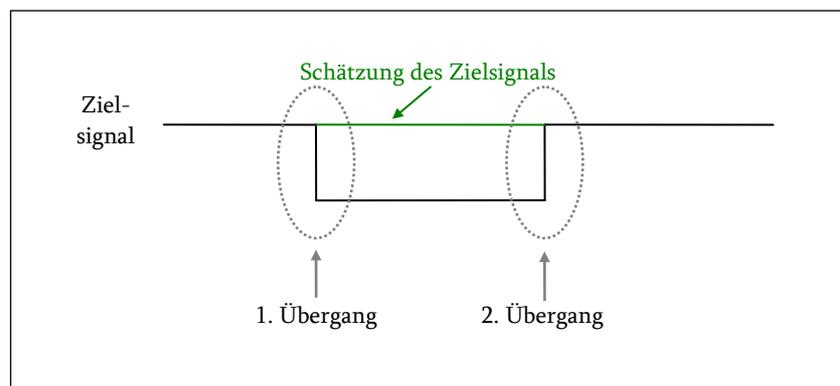


Abb. 2-2: Kritische Stellen der Ausfallsverschleierung

In Abb. 2-2 sind die zwei kritischen Stellen aufgezeigt, die bei der Erzeugung des geschätzten Zielsignals speziell behandelt werden müssen. Es handelt sich dabei um die Übergänge vom Zielsignal auf dessen Schätzung und umgekehrt. Die einfachste Realisierung dieser Übergänge wäre eine lineare Überblendung zwischen beiden Signalen. An der ersten Stelle ist dies jedoch aus folgendem Grund nicht möglich: der Beginn eines Ausfalls wird durch den ersten, fehlenden Block gekennzeichnet, das bedeutet, dass kein Zielsignal im aktuellen sowie dem/den folgenden Block/Blöcken für die Überblendung verfügbar ist. Eine "zeitliche Verlagerung" der Überblendung in die Vergangenheit ist ebenfalls nicht durchführbar, da zu dem Zeitpunkt, zu dem die dafür benötigten Samples aktuell sind, noch keine Information über den späteren Ausfall vorliegt. Die Idee, das Signal zu verzögern, damit ein Ausfall schon

im Vorhinein erkannt werden kann, wird wegen der dadurch verursachten zusätzlichen Latenzzeit nicht in Betracht gezogen.

Somit muss eine Technik gefunden werden, die ohne Überblendung einen stetigen Übergang zwischen den beiden Signalen gewährleistet. Die in dieser Arbeit untersuchten signaltheoretischen Methoden werden speziell auf diese Forderung hin untersucht.

Die Behandlung des zweiten Übergangs – vom geschätzten Zielsignal zum originalen – ist weniger kritisch. Hier kann eine Überblendung realisiert werden, da es möglich ist, die Länge des geschätzten Zielsignals darauf abzustimmen. Dies ist in Abb. 2-3 schematisch dargestellt.

Eine Alternative zur linearen Überblendung bieten Interpolationsfunktionen wie beispielsweise *spline*- oder Bezier-Kurven [Bartels 2006]. Die Verwendung dieser Methoden ermöglicht eine Verkleinerung des Übergangsbereiches, die beim ersten Übergang von Vorteil ist, wenn die Ausfallsverschleierung als Kombination von Extrapolation und Substitution realisiert wird, da dadurch der Extrapolationsbereich minimiert werden kann.

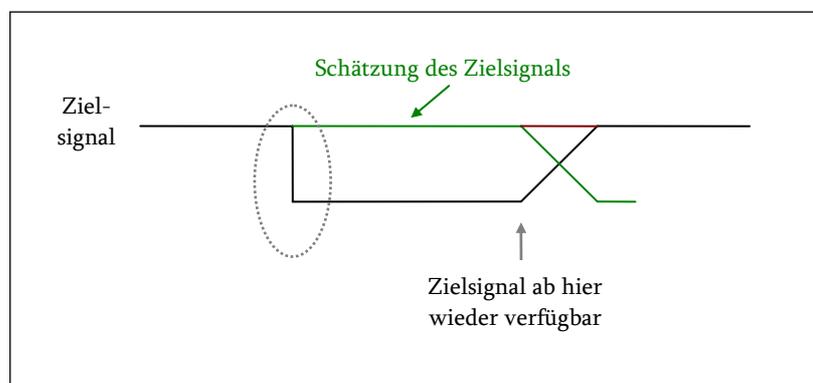


Abb. 2-3: Übergang von der Schätzung auf das originale Zielsignal

Zusammengefasst erfüllen die vier Hauptzweige folgende Aufgaben:

- (1) *first DO frame* – Start der Signalerzeugung; Schätzung und Speicherung der dafür notwendigen Parameter,
- (2) *middle DO frame* – Fortführung der Signalerzeugung,
- (3) *first original frame after DO* – Erzeugung des zum Überblenden notwendigen Signals; Start der Überblendung,

(4) *no DO (frame correct)* – ggf. Weiterführung der Überblendung bzw. "Durchschleifen" des Zielsignals.

In den folgenden Abschnitten werden diese im Detail erläutert sowie konkrete Verschleierungstechniken vorgestellt und diskutiert, die zur Parameterschätzung und zur Schätzung des Zielsignals verwendet werden. Den Schwerpunkt bilden dabei Extrapolationstechniken, die auf dem Prinzip der linearen Prädiktion beruhen. Eine vielversprechende Alternative stellt die Substitution durch passende Ausschnitte aus der Signalhistorie dar. Eine Verknüpfung beider Methoden führt zu einer speziellen Modifikation der von Vaseghi [Vaseghi 1996] vorgeschlagenen und in [Etter 1996] ebenfalls erwähnten statistischen Interpolation. Diese dient unter anderem als Basis für die Realisierung der Pufferverbesserung im Modul "*improve buffer*".

2.2. Signaltheoretische Grundlagen

Dieser Abschnitt beinhaltet eine Zusammenfassung der theoretischen Grundlagen der Signalverarbeitung, die als Basis der vorliegenden Dissertation dienen. Da sich die gesamte Arbeit ausschließlich mit der Verarbeitung von digitalen Signalen beschäftigt, wird auf die Beziehung zu zeitkontinuierlichen Funktionen kaum eingegangen. Das vorgeschlagene Verfahren ist grundsätzlich durch (a) die Betrachtung und Verarbeitung der (digitalen) Signale im Zeitbereich und (b) die Beschreibung der zeitdiskreten Signalfolgen mithilfe von parametrischen Modellen geprägt.

2.2.1. Deterministische und stochastische Signale

Ein eindimensionales Signal beschreibt die Abhängigkeit einer physikalischen Größe von einer zweiten. Dieser Zusammenhang wird mathematisch als Funktion einer Variable

bezeichnet: $y = Fkt\{x\}$. Sowohl Funktion als auch Variable können kontinuierliche oder diskrete Größen sein. Bei einem digitalen Signal sind beide Größen diskret.

Eine der wichtigsten Repräsentationen in der Signalverarbeitung sind zeitdiskrete Signale, die eine (geordnete) Folge von Signalpunkten entlang der Zeitachse definieren. Im Allgemeinen erfolgt die Abtastung in regelmäßigen Abständen, d. h. die Punkte sind gleichmäßig entlang der Zeitachse angeordnet, wobei jeder Abtastpunkt nT den Funktionswert zu diesem Zeitpunkt beinhaltet. Der Parameter T definiert das (physikalische) Zeitintervall zwischen zwei Abtastpunkten in Sekunden, n ist eine ganze Zahl $n \in \mathbb{Z}$ und dient in der diskreten Domäne als "Einheit" der Zeitachse.

Ein deterministisches Signal kann vollständig und exakt durch eine mathematische Formel beschrieben werden, z. B. Kosinusfunktion: $s(n) = A \cos(2\pi\nu n)$, wobei A die (diskrete) Amplitude und ν die (diskrete) Frequenz der Funktion definieren. Beide Parameter können entweder konstante oder wiederum zeitlich veränderliche Größen sein, z. B. $s(n) = A(n) \cos[2\pi\nu(n)n]$. Die Superposition solcher Einzelschwingungen erlaubt die Modellierung komplexer Signale: $s(n) = \sum_{l=0}^{L-1} A_l(n) \cos[2\pi\nu_l(n)n]$. Eine weitere wichtige Eigenschaft deterministischer Signale, die sich aus der exakten Beschreibbarkeit ergibt, ist die Fähigkeit, die Zukunft eines Signals eindeutig vorherzubestimmen, d. h. ein Signal ist zu jedem beliebigen Zeitpunkt vollständig prädizierbar.

Im Gegensatz dazu können die Werte eines nichtdeterministischen Signals nicht vollständig vorhergesagt werden. Die Funktion kann zum Zeitpunkt n einen zufälligen Wert aus einer (endlichen) Werteschar annehmen. Es ist somit nicht möglich, ein Zufallssignal mithilfe einer eindeutigen geschlossenen mathematischen Formel zu beschreiben. Zur Charakterisierung der statistischen Eigenschaften verwendet man stochastische Größen, wie Verteilungs- und Dichtefunktionen, Mittelwert, Varianz, etc.

Die grundlegende Idee, eine mathematische Formulierung für ein nichtdeterministisches Signal zu entwickeln, liegt darin, einen Zufallsprozess als Modell für diese Signalklasse zu

verwenden.⁴ Eine Reihe von zufälligen Zahlen wird generiert, indem das Ergebnis einer beliebigen Operation (eines Prozesses) indiziert und notiert wird: $\{x(n)\}$ beschreibt die sich dadurch ergebende Zufallsfolge. Bei zeitdiskreten Signalen wird der Index n der physikalischen Größe Zeit zugeordnet, allgemein kann eine beliebige physikalische Dimension dafür verwendet werden. Aus dieser Definition folgt, dass der n -te Wert der Folge, $x(n)$, als ein spezieller Wert einer Zufallsvariable x_n angesehen werden kann. Die gesamte Menge von Zufallsvariablen $\{x_n\}$ definiert zusammen mit deren einzelnen Wahrscheinlichkeiten den Zufallsprozess.

Die Zufallsfolge $\{x(n)\}$ wird auch als Realisation oder Stichprobenfolge des Prozesses bezeichnet. Aus ihrer Definition kann abgeleitet werden, dass theoretisch unendlich viele, unterschiedliche Realisationen generiert werden können. Die Gesamtheit aller Folgen stellt in der Stochastik eine wichtige Größe dar.

Um einen Zufallsprozess als Modell für ein zeitdiskretes Signal einsetzen zu können, betrachtet man nun eine spezielle Zufallsfolge aus der Menge aller (praktisch) möglichen, welche genau diesen Prozess charakterisieren. Anders interpretiert dient der Zufallsprozess als Repräsentation eines nichtdeterministischen Signals. Im Folgenden wird davon ausgegangen, dass das Signal verfügbar ist, jedoch Struktur bzw. Eigenschaften des Prozesses unbekannt und deshalb aus dem Signal zu schätzen sind.

2.2.2. Beschreibung eines Zufallsprozesses

Die Wahrscheinlichkeit, dass die Zufallsvariable x_n kleiner oder gleich einem bestimmten Wert \mathbf{x}_n aus der Menge $\{x_n\}$ ist, wird durch die Verteilungsfunktion ausgedrückt:

$$P_{x_n}(\mathbf{x}_n, n) = \text{Wahrscheinlichkeit}(x_n \leq \mathbf{x}_n).$$

Steht x_n ein kontinuierlicher Wertebereich zur Verfügung, ergibt sich die Dichtefunktion

als Ableitung von $P_{x_n}(\mathbf{x}_n, n)$ nach \mathbf{x}_n : $p_{x_n}(\mathbf{x}_n, n) = \frac{\partial P_{x_n}(\mathbf{x}_n, n)}{\partial \mathbf{x}_n}$. Umgekehrt gilt:

⁴ Formulierungen und Definitionen werden in Anlehnung an Kapitel 8 aus [Oppenheim 1975] verwendet.

$P_{x_n}(\mathbf{x}_n, n) = \int_{-\infty}^{\mathbf{x}_n} p_{x_n}(x, n) dx$. Bei einem diskreten Wertebereich wird alternativ zur

Dichtefunktion die Massenfunktion definiert, die die Wahrscheinlichkeit angibt, bei der die Zufallsvariable x_n exakt einen Wert \mathbf{x}_n annimmt: $p_{x_n}(\mathbf{x}_n, n) = \text{Wahrscheinlichkeit}(x_n = \mathbf{x}_n)$.

Damit ergibt sich folgender Zusammenhang zwischen Verteilung und Massenfunktion:

$$P_{x_n}(\mathbf{x}_n, n) = \text{Wahrscheinlichkeit}(x_n \leq \mathbf{x}_n) = \sum_{x \leq \mathbf{x}_n} p_{x_n}(x, n).$$

Die Einführung von Durchschnittswerten oder Mittelungen bietet weitere nützliche Möglichkeiten, bestimmte Eigenschaften von Zufallsvariablen zu extrahieren. Der

Mittelwert eines Prozesses ist definiert als: $\mu_{x_n} = E\{x_n\} = \int_{-\infty}^{\infty} xp_{x_n}(x, n) dx$, wobei $E\{*\}$ den

Erwartungswert einer Variable (oder Funktion) kennzeichnet.⁵ Weiters kann mithilfe des Erwartungswertes der Mittelwert der Funktion einer Zufallsvariable – der Zufallsfolge

$$g(x_n) - \text{bestimmt werden: } E\{g(x_n)\} = \int_{-\infty}^{\infty} g(x)p_{x_n}(x, n) dx.$$

Den quadratischen Mittelwert einer Variable erhält man mit $E\{(x_n)^2\} = \int_{-\infty}^{\infty} (x)^2 p_{x_n}(x, n) dx$,

die Varianz definiert den quadratischen Wert der vom Mittelwert befreiten Variable:

$$E\{(x_n - \mu_{x_n})^2\} = E\{(x_n)^2\} - \mu_{x_n}^2 = \sigma_{x_n}^2.$$

Wird x_n quantisiert (diskreter Wertebereich), muss die Integration durch die Summation

$$\text{angenähert werden, z. B. } E\{g(x_n)\} = \sum_x g(x)p_{x_n}(x, n).$$

Korrelations- und Kovarianzfolgen werden ebenfalls zur Durchschnittsberechnung verwendet, wobei man folgende Arten unterscheidet:

⁵ Der Erwartungswert $E\{*\}$ ist eine Grundfunktion in der Stochastik [Kreyszig 1993], Rechenregeln und Eigenschaften des Operators in Bezug auf signaltheoretische Anwendungen finden sich u. a. in [Oppenheim 1975], [Deller 2000], etc.

$$\text{Autokorrelation: } \phi_{xx}(n, m) = E \left\{ x_n x_m^* \right\} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mathbf{x}_n \mathbf{x}_m^* p_{x_n, x_m}(\mathbf{x}_n, n, \mathbf{x}_m, m) d\mathbf{x}_n d\mathbf{x}_m,$$

$$\text{Kreuzkorrelation: } \phi_{xy}(n, m) = E \left\{ x_n y_m^* \right\} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy^* p_{x_n, y_m}(x, n, y, m) dx dy,$$

$$\text{Autokovarianz: } \gamma_{xx}(n, m) = E \left\{ (x_n - \mu_{x_n})(x_m - \mu_{x_m})^* \right\} = \phi_{xx}(n, m) - \mu_{x_n} \mu_{x_m},$$

$$\text{Kreuzkovarianz: } \gamma_{xy}(n, m) = E \left\{ (x_n - \mu_{x_n})(y_m - \mu_{y_m})^* \right\} = \phi_{xy}(n, m) - \mu_{x_n} \mu_{y_m}.$$

Die Autokorrelation ist ein Maß für die gegenseitige Abhängigkeit der Werte eines Zufallsprozesses zu unterschiedlichen Zeitpunkten n bzw. m und wird daher zur Beschreibung der zeitlichen Varianz eines Zufallssignals verwendet. In gleicher Weise interpretiert die Kreuzkorrelation die Beziehung zweier unterschiedlicher Zufallssignale. Mittelwert, quadratischer Mittelwert und Varianz gehören zu einer Klasse von Mittelungsverfahren, die keine zeitliche Information beinhalten; für stationäre Prozesse liefern sie konstante Werte. Folglich weisen Korrelations- und Kovarianzanalysen im Allgemeinen einen höheren Informationsgehalt auf.

Die Idee, nichtdeterministische Signale als (theoretisch unendliche) Menge von Realisationen eines Zufallsprozesses zu interpretieren, ist zwar für die Definition der obigen Mittelungen zweckdienlich, jedoch beim Umgang mit realen Signalen unbrauchbar. Anstatt die Menge der Realisationen zu betrachten, wäre es hilfreicher, eine spezielle Folge daraus zu wählen und die statistischen Eigenschaften lediglich von dieser einen abzuleiten. Die Umsetzung des Gedankens führt zur Definition der zeitlichen Mittelung eines

Zufallsprozesses: $\langle x_n \rangle = \lim_{N \rightarrow \infty} \frac{1}{2N+1} \sum_{n=-N}^N x_n$ und der zeitlichen Autokorrelationsfolge (AKF):

$\langle x_n x_{n+m}^* \rangle = \lim_{n \rightarrow \infty} \frac{1}{2N+1} \sum_{n=-N}^N x_n x_{n+m}^*$. Für stationäre Prozesse mit endlichem Mittelwert existieren

beide Grenzwerte (siehe [Oppenheim 1975], Kap. 8.2.2). Die Bedingung für die Gleichsetzung der zeitlichen Mittelungen mit dem Scharmittelwert wird Ergodizität genannt. Demnach entspricht der zeitliche Mittelwert eines ergodischen Prozesses dem

Erwartungsmittelwert, $\langle x(n) \rangle = E\{x_n\}$ und die zeitliche AKF der Autokorrelationsfolge der Menge der Zufallsvariablen, $\langle x(n)x(n+m)^* \rangle = E\{x_n x_{n+m}^*\} = \phi_{xx}(m)$. In den meisten Problemstellungen der Signalverarbeitung – wie z. B. im vorliegenden Fall – wird die Annahme getroffen, dass sich die Klasse der Eingangssignale auf ergodische Funktionen beschränkt. Streng genommen kann die Bedingung aufgrund der unendlichen zeitlichen Ausdehnung nicht erfüllt werden, weshalb in der Praxis Schätzwerte berechnet werden.

2.3. Prädiktion

In der SV werden mit dem Begriff "Prädiktion" jene Methoden der Signalfortführung zusammengefasst, bei denen ein Signal über einen bestimmten Zeitraum beobachtet wird, um charakteristische Parameter desselben abzuleiten. In einer anschließenden Verarbeitungsstufe (i. A. in der Resynthese) dienen diese dazu, die Zukunft des Signals vorherzusagen bzw. zu präzisieren. Das Extrahieren der Eigenschaften führt zur Erstellung eines Modells, mit dem das Signal in die Zukunft fortgesetzt wird. Bei der Restauration von digitalisierten Analogaufnahmen spielen Prädiktionsalgorithmen eine wichtige Rolle, da im Allgemeinen das Signal gespeichert ist und somit die gesamten Daten gleichzeitig und zu Beginn der Verarbeitung verfügbar sind.

2.3.1. Parametrische Modelle für nichtdeterministische Signale

Ein nichtdeterministisches Signal, das eine spezielle Realisation eines Zufallsprozesses ist, kann mithilfe dieses Zufallsprozesses modelliert werden. Anders formuliert bildet ein Zufallsprozess das Modell für eine diskrete Zeitfolge. Ähnlich wie bei deterministischen Signalen sind auch für die Beschreibung von Modellen stochastischer Signale mehrere

Parameter notwendig. Zur Schätzung dieser Parameter sind in der Signalverarbeitung drei Methoden gebräuchlich: autoregressive (AR), gleitende Mittelwerte (MA) und – als Kombination der beiden – ARMA Modelle. Entsprechend ihrer Repräsentation mittels Differenzgleichung sind auch die Bezeichnungen Polstellen-, Nullstellen- sowie Pol-Nullstellen-Filter gebräuchlich:

$$\text{Polstellen-Filter: } x(n) = -\sum_{k=1}^p a(k)x(n-k) + u(n) \quad (2.1)$$

$$\text{Nullstellen-Filter: } x(n) = \sum_{k=0}^o b(k)u(n-k), \text{ mit } b(0) = 1 \quad (2.2)$$

$$\text{Pol-Nullstellen-Filter: } x(n) = -\sum_{k=1}^p a(k)x(n-k) + \sum_{k=0}^o b(k)u(n-k) \quad (2.3)$$

Das Ausgangssignal des jeweiligen Filters wird $x(n)$ bezeichnet, $u(n)$ ist dessen Eingang und $a(k)$ bzw. $b(k)$ spezifizieren die Modellparameter. Wird als Eingangsfolge Weißes Rauschen verwendet, so lässt sich der Ausgang vollständig durch die Modellparameter und die Varianz des Weißen Rauschens bestimmen.⁶ Im Folgenden wird gezeigt, wie aus der Autokorrelationsfolge von $x(n)$ die Parameter und die Varianz geschätzt werden. Grundsätzlich kann jedes der drei Modelle dafür verwendet werden, wobei das Polstellen-Modell als einziges einen linearen Zusammenhang zwischen AKF und Modellparameter aufweist und somit einen entscheidenden Vorteil gegenüber den beiden anderen Methoden besitzt. Weiters existieren eine Reihe von recheneffizienten, schnellen Implementierungen von AR-Modellen, dies wiederum ist eines der Hauptkriterien in der Zieldefinition der vorliegenden Arbeit. Da MA- und ARMA-Modelle nicht als Grundlage für die weiteren Entwicklungen von DROCON ausschlaggebend sind, wird nicht näher darauf eingegangen. Details finden sich unter anderem in [Marple 1987], [Haykin 2002], [Vaseghi 1996].

⁶ Als Weißes Rauschen bezeichnet man einen Zufallsprozess mit unkorrelierten Zufallsvariablen, dessen Mittelwert ist $\mu_{WR} = E\{u(n)\} = 0$ und die Varianz ist konstant mit $\sigma_{WR}^2 = konst$.

2.3.2. Beziehung zwischen AR-Modellparameter und AKF

Die mathematische Beziehung erhält man durch Multiplikation von Gl. (2.1) mit $x^*(n-m)$ und Bildung des Erwartungswertes:

$$E\{x(n)x^*(n-m)\} = -\sum_{k=1}^p a(k)E\{x(n-k)x^*(n-m)\} + E\{u(n)x^*(n-m)\} \quad (2.4)$$

Die linke Seite der Gleichung entspricht der Autokorrelation $\phi_{xx}(m)$, der erste Erwartungswert auf der rechten Seite $-E\{x(n-k)x^*(n-m)\}$ entspricht der Autokorrelation mit dem Versatz $(m-k) - \phi_{xx}(m-k)$. Der zweite Erwartungswert $-E\{u(n)x^*(n-m)\}$ repräsentiert die Kreuzkorrelation $-\phi_{ux}(m)$, für die aufgrund der Definition von $u(n)$ als unkorrelierter Zufallsprozess gilt: $\phi_{ux}(m) = \sigma_u^2$ für $m=0$, bzw. $\phi_{ux}(m) = 0$ für $m > 0$. Mithilfe dieser Formulierungen ergibt sich folgender Zusammenhang zwischen dem AR-Modell und der AKF:

$$\phi_{xx}(m) = \begin{cases} -\sum_{k=1}^p a(k)\phi_{xx}(m-k) & m > 0 \\ -\sum_{k=1}^p a(k)\phi_{xx}(-k) + \sigma_u^2 & m = 0 \\ \phi_{xx}^*(-m) & m < 0 \end{cases} \quad (2.5)$$

Die Verwendung der Matrixschreibweise führt zur Definition der sogenannten Yule-Walker-Gleichungen:

$$\begin{pmatrix} \phi_{xx}(0) & \phi_{xx}(-1) & \cdots & \phi_{xx}(-p) \\ \phi_{xx}(1) & \phi_{xx}(0) & \cdots & \phi_{xx}(-p+1) \\ \vdots & \vdots & \ddots & \vdots \\ \phi_{xx}(p) & \phi_{xx}(p-1) & \cdots & \phi_{xx}(0) \end{pmatrix} \begin{pmatrix} 1 \\ a(1) \\ \vdots \\ a(p) \end{pmatrix} = \begin{pmatrix} \sigma_u^2 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \quad (2.6)$$

Ist eine Schätzung der AKF der Ausgangsfolge bekannt, so können dessen Modellparameter und die Varianz des Weißen Rauschens am Eingang geschätzt werden.⁷ Man spricht in diesem Fall von der Analyse des Zufallsprozesses.

Analog dazu lässt sich die Synthese bzw. Generation einer Signalfolge definieren, indem Weißes Rauschen mit bekannter Varianz als Eingangssignal eines durch seine Parameter gekennzeichneten Filters verwendet wird. Als Ausgangsfolge erhält man ein gefiltertes Signal, welches mithilfe des Prozessmodells generiert wurde. Letztere Definition erlaubt direkt die Anwendung des Signalmodells nach Gl. (2.1) über eine Länge von N Samples:⁸

$$\begin{pmatrix} x(1) \\ x(2) \\ \vdots \\ x(N) \end{pmatrix} = - \begin{pmatrix} x(0) & x(-1) & \dots & x(-p+1) \\ x(1) & x(0) & \dots & x(-p+2) \\ \vdots & \vdots & \ddots & \vdots \\ x(N-1) & x(N-2) & \dots & x(-p+N) \end{pmatrix} \begin{pmatrix} a_a(1) \\ a_a(2) \\ \vdots \\ a_a(p) \end{pmatrix} + \begin{pmatrix} u_{\sigma,a}(1) \\ u_{\sigma,a}(2) \\ \vdots \\ u_{\sigma,a}(N) \end{pmatrix} \quad (2.7)$$

Für die Implementierung im E-DROCON-Algorithmus sind beide Interpretationen interessant. Zum einen wird das Signal, solange kein Ausfall stattfindet, als Ausgang eines AR-Prozesses betrachtet. Mithilfe der in einem Puffer gespeicherten letzten Signalsamples werden die Modellparameter $[a_a(1) \ a_a(2) \ \dots \ a_a(p)]^T$ und die Varianz des Weißen Rauschens σ_u^2 geschätzt. Bei Signalausfall erfolgt die Generation des Signals, das zur Überbrückung der fehlenden Samples dient, wobei Weißes Rauschen der zuvor bestimmten Varianz als Anregungssignal angenommen wird, siehe Gl. (2.7).

In der Signalverarbeitung trifft man häufig auf Anwendungen, in denen Analyse- und Syntheseeinheiten in dieser Weise nacheinander geschaltet sind. Allerdings sind beide Teile meist örtlich voneinander getrennt, weshalb der Vorteil des Gesamtsystems unter anderem darin liegt, dass anstatt eines vollbandigen Audiosignals lediglich einige Parameter übertragen werden müssen. Dadurch kann die Bandbreite der Übertragungskanäle reduziert

⁷ Eine eindeutige Bestimmung von $a(k)$ und σ_u^2 ist nicht möglich, da die AKF aufgrund ihrer Definition selbst eine Schätzung darstellt.

⁸ Das tiefgestellte a kennzeichnet die Werte, die aus der Signalanalyse ermittelt wurden.

und die Geschwindigkeit des Transfers erhöht werden. In der vorliegenden Systemdefinition liegen Analyse und Synthese unmittelbar hintereinander – dazwischen ist keine physikalische Übertragungsstrecke notwendig. Eine gleichzeitige Verarbeitung des Audiosignals aus dem Puffer (z. B. in Form einer direkten Substitution von Samples aus dem Puffer) wird so möglich.⁹ Der Vorteil der Modell-Realisation gegenüber einer "einfachen" Substitution besteht in der Tatsache, dass bei langen Ausfällen, die sich jedoch innerhalb der Kurzzeitstationarität des Signals befinden, die Qualität der Verschleierung durch Autoregression besser ist, [Herre 1995], [Synnes 1999], [Wah 2000a]. Ein Kriterium, das bei der Implementierung der Methode in dieser Form im E-DROCON-Algorithmus dennoch zu berücksichtigen ist, liegt im un stetigen Übergang von Zielsignal zu Extrapolation.

Eine alternative Interpretation des AR-Modells führt zur Definition des linearen Vorwärtsprädiktionsfehlerfilters (LVPPF). Im Folgenden wird gezeigt, dass beide Ansätze mathematisch äquivalent sind, aber aus unterschiedlichen Themengebieten hergeleitet werden. Die Modellierung eines AR-Prozesses hat ihren Ursprung in der Stochastik, wohingegen die Prädiktionsfilter und deren Modifikationen der Signaltheorie entstammen, [Makhoul 1975].

2.3.3. Lineare Prädiktion (LP)

Bei der Analyse liegt das Bestreben darin, einen oder mehrere Abtastpunkte des zukünftigen Signals vorherzusagen, d. h. es wird versucht, mit einer gegebenen Menge von vergangenen Signalsamples $\{x(n-1) \ x(n-2) \ \dots \ x(n-q)\}$ das Signal zum Zeitpunkt n zu ermitteln.

Die lineare Prädiktion ermöglicht diese Art der schrittweisen Vorhersage der Zukunft des Signals, indem der jeweils zu prädizierende Wert als Linearkombination seiner unmittelbar vorangegangenen Samples betrachtet wird. Aus dieser Interpretation leitet sich die Definition der linearen Vorwärtsprädiktion ab:

⁹ Eine Technik, die diesen Ansatz untersucht, ist im nachfolgenden Abschnitt angeführt.

$$\hat{x}(n) = \sum_{k=1}^q w_f(k) x(n-k) \quad (2.8)$$

Zur Darstellung der Vorwärtsprädiktion wird ein lineares Transversalfilter verwendet, siehe Abb. 2-4, worin die Gewichte der einzelnen Filterzweige mit $w_f(k)$ ($0 \leq k \leq q$) gekennzeichnet sind und die Prädiktionskoeffizienten repräsentieren. Eine übliche Methode der Gewinnung optimaler Koeffizienten lässt sich über die Minimierung des mittleren Quadrats des Prädiktionsfehlers realisieren, der definitionsgemäß aus der Differenz von wahrem Wert und Schätzwert gebildet wird:

$$e(n) = x(n) - \hat{x}(n) = x(n) - \sum_{k=1}^q w_f(k) x(n-k) \quad (2.9)$$

Damit ergibt sich der mittlere quadratische Fehler:

$$\begin{aligned} E\{e^2(n)\} &= E\left\{\left[x(n) - \sum_{k=1}^q w_f(k) x(n-k)\right]^2\right\} = \\ &= E\{x^2(n)\} - 2 \sum_{k=1}^q w_f(k) E\{x(n)x(n-k)\} + \sum_{k=1}^q w_f(k) \sum_{j=1}^q w_f(j) E\{x(n-k)x(n-j)\} = \\ &= \phi_{xx}(0) - 2\boldsymbol{\phi}^H \mathbf{w}_f + \mathbf{w}_f^H \boldsymbol{\Phi} \mathbf{w}_f \end{aligned}$$

In der letzten Zeile wird der Vektor der Filtergewichte $\mathbf{w}_f = [w_f(1) \ w_f(2) \ \dots \ w_f(q)]^T$ eingeführt, weiters werden unter Zuhilfenahme des Signalvektors $\mathbf{x}_{n-1} = [x(n-1) \ x(n-2) \ \dots \ x(n-q)]^T$ die Autokorrelationsmatrix $\boldsymbol{\Phi}$ und der Korrelationsvektor $\boldsymbol{\phi}$ definiert:

$$\boldsymbol{\Phi} = E\{\mathbf{x}_{n-1} \mathbf{x}_{n-1}^H\} = \begin{pmatrix} \phi_{xx}(0) & \phi_{xx}(-1) & \dots & \phi_{xx}(-q+1) \\ \phi_{xx}(1) & \phi_{xx}(0) & \dots & \phi_{xx}(-q+2) \\ \vdots & \vdots & \ddots & \vdots \\ \phi_{xx}(q-1) & \phi_{xx}(q-2) & \dots & \phi_{xx}(0) \end{pmatrix},$$

$$\boldsymbol{\phi} = E\{\mathbf{x}_{n-1} x^*(n)\} = [\phi_{xx}(-1) \ \phi_{xx}(-2) \ \dots \ \phi_{xx}(-q)]^T.$$

Der Ableitung des mittleren quadratischen Fehlers nach dem Vektor der Filtergewichte ergibt $\frac{\partial}{\partial \mathbf{w}_f} E\{e^2(n)\} = -2\boldsymbol{\phi}^H + 2\mathbf{w}_f^H \boldsymbol{\Phi}$. Durch Nullsetzen des Gradienten erhält man ein

Gleichungssystem, welches zur Beschreibung der Vorwärtsprädiktion herangezogen wird:

$$\boldsymbol{\Phi} \mathbf{w}_f = \boldsymbol{\phi}. \quad (2.10)$$

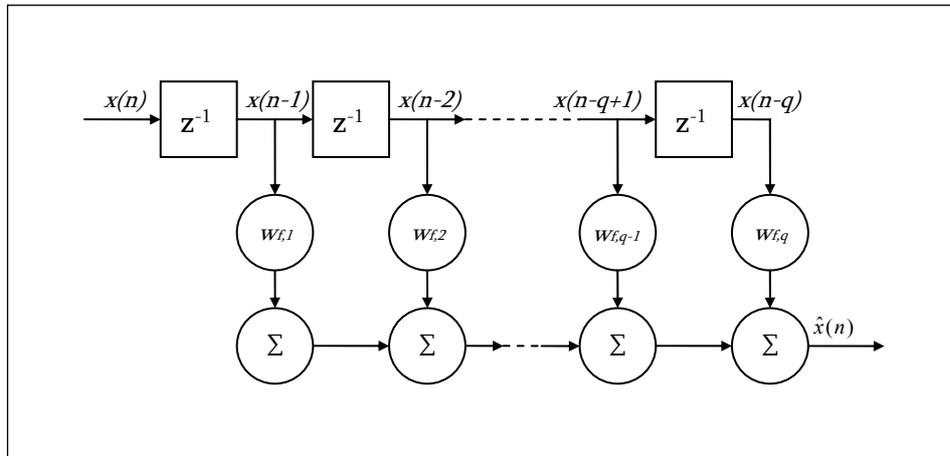


Abb. 2-4: Lineares Transversalfilter, Vorwärtsprädiktion

Mit den daraus resultierenden optimalen Filtergewichten $\mathbf{w}_f = \boldsymbol{\Phi}^{-1} \boldsymbol{\phi}$ wird die minimierte mittlere Fehlerleistung zu

$$E\{e^2(n)\} = P_q = \phi_{xx}(0) - \boldsymbol{\phi}^H \mathbf{w}_f. \quad (2.11)$$

Zusammen werden Gl. (2.10) und Gl. (2.11) als erweiterte Wiener-Hopf-Gleichungen bezeichnet:

$$\begin{pmatrix} \phi_{xx}(0) & \phi_{xx}(-1) & \dots & \phi_{xx}(-q) \\ \phi_{xx}(1) & \phi_{xx}(0) & \dots & \phi_{xx}(-q+1) \\ \vdots & \vdots & \ddots & \vdots \\ \phi_{xx}(q) & \phi_{xx}(q-1) & \dots & \phi_{xx}(0) \end{pmatrix} \begin{pmatrix} 1 \\ -w_f(1) \\ \vdots \\ -w_f(q) \end{pmatrix} = \begin{pmatrix} P_q \\ 0 \\ \vdots \\ 0 \end{pmatrix} \quad (2.12)$$

Wird der Fehler $e(n)$ bzw. der minimierte quadratische Fehler P_q in das Gleichungssystem miteinbezogen, nennt man das daraus entstehende Filter lineares

Vorwärtsprädiktionsfehlerfilter mit den Koeffizienten $\mathbf{c} = [1 \quad -w_f(1) \quad \dots \quad -w_f(q)]^T$. Ein Vergleich von Gl. (2.12) mit Gl. (2.6) zeigt, dass beide Systeme exakt dieselbe mathematische Struktur besitzen, dennoch ist deren signaltheoretische Interpretation unterschiedlich. Nur unter der Voraussetzung, dass das Eingangssignal des LVPFF als autoregressiver Prozess generiert wurde, erhält man am Ausgang eine unkorrelierte Zufallsfolge mit den gleichen Eigenschaften wie ein Weißer Zufallsprozess (d. h. mittelwertfrei, mit konstanter Varianz und Weißem Leistungsdichtespektrum, vgl. Definition "Weißer Zufallsprozess" bzw. "Weißes Rauschen"). In diesem Fall sind die AR-Parameter und die LP-Koeffizienten ident. Allgemein gilt jedoch lediglich, dass der Fehler des LVPFF mit der Schätzung des Prädiktors $\hat{x}(n)$ nicht korreliert, was nicht notwendigerweise bedeutet, dass er ein Weißer Zufallsprozess sein muss.

Bei einem AR-Prozess stellt das Signal $x(n)$ den Ausgang eines Zufallsprozesses dar, dessen Eingang Weißes Rauschen $u(n)$ ist. Das LVPFF hat im Gegensatz dazu als Eingangsfolge das Signal $x(n)$, die Ausgangsfolge bildet der Prädiktionsfehler $e(n)$. Beide Strukturen sind in Abb. 2-5 dargestellt. Es zeigt sich, dass das LVPFF einem reinen Nullstellenfilter entspricht, wohingegen das AR-Modell ein reines Polstellenfilter repräsentiert. Folgende zwei Überlegungen lassen sich daraus ableiten:

- (1) Die Anwendung des LVPF-Filters auf einen stationären Prozess $x(n)$ kann als Analyse betrachtet werden, d. h. man sucht nach einer parametrischen Beschreibung für $x(n)$. Bei geeigneter Wahl der Prädiktionsordnung q wird der Fehler $e(n)$ zu einem völlig unkorrelierten Prozess. Aufgrund dieser Eigenschaft wird er auch als Residuum bezeichnet und dient zusammen mit den Koeffizienten $c(k)$ zur Beschreibung von $x(n)$.
- (2) Obwohl die Anwendung des AR-Modells ebenfalls zur Beschreibung eines stochastischen Prozesses durch die Modellparameter dient, wird dies hier als Synthese interpretiert. Den Grund dafür liefert die Betrachtung des AR-Prozesses als reines Polstellenfilter, dessen

Koeffizienten ausschließlich auf verzögerte Samples des Filterausgangs wirken, [Haykin 2002].

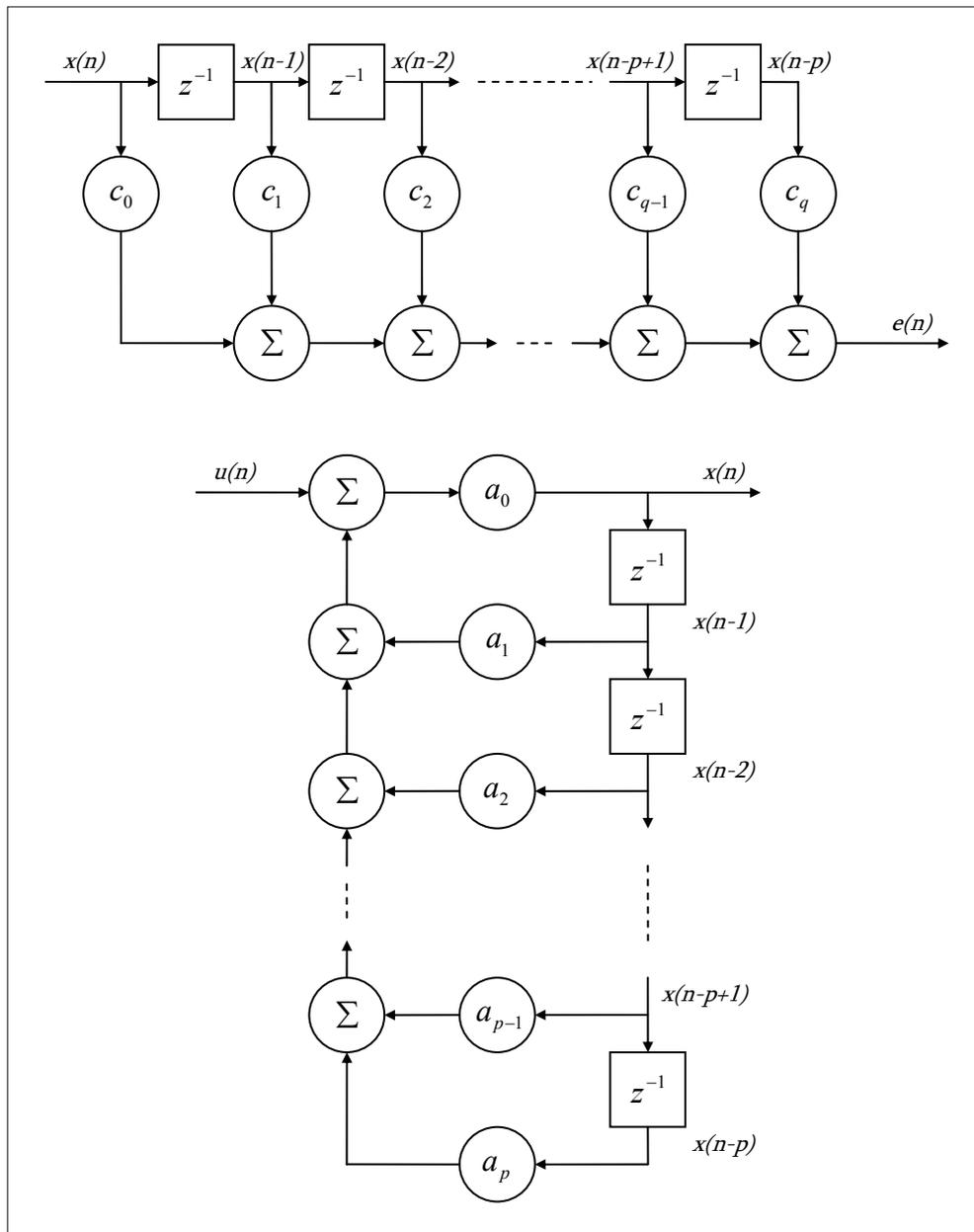


Abb. 2-5: LVPFF und AR-Modell

Aufgrund der aufgestellten Modelle für AR und LP lässt sich erkennen, dass zur Ermittlung der gesuchten Größen – es sind dies die AR-Parameter $a(k)$ bzw. LP-Koeffizienten $c(k)$ – immer eine Matrixinversion notwendig ist. Dies ist mit einem nicht vernachlässigbaren Rechenaufwand verbunden, was zur Notwendigkeit führt, effiziente und schnelle Algorithmen dafür einzusetzen. Der Levinson-Durbin-Algorithmus, der die Toeplitz-

Struktur der Autokorrelationsmatrix ausnützt, bietet eine verbesserte Möglichkeit im Vergleich zu Standardtechniken wie z. B. dem Gaußschen Eliminationsverfahren. Zur Berechnung wird zusätzlich die lineare Rückwärtsprädiktion benötigt, bei der mit der Menge $\{x(n) \ x(n-1) \ \dots \ x(n-q+1)\}$ das Sample $x(n-q)$ geschätzt wird:

$$\hat{x}(n-q) = \sum_{k=1}^q w_b(k) x(n-q+k) \quad (2.13)$$

Der Zusammenhang der Koeffizienten von Vor- und Rückwärtsprädiktion, $w_f(k)$ und $w_b(k)$, ist definiert durch: $w_b(k) = w_f^*(q-k)$. Die optimalen Koeffizienten $c(k)$ des Prädiktionsfehlerfilters werden letztlich rekursiv geschätzt, eine Zusammenfassung der Berechnungsschritte findet sich in [Vaseghi 1996], Seite 198.

Die Implementierung der linearen Prädiktion erfolgt im E-DROCON-Verfahren über einen zweistufigen Algorithmus. In der ersten Stufe – der Analyse – werden mit der verfügbaren Signalfolge \mathbf{x}_{n-1} die Filterkoeffizienten geschätzt. Die Rekonstruktion des Zielsignals $x(n)$ findet nachfolgend in einer separaten Stufe statt, in der die zuvor ermittelten Koeffizienten in Gl. (2.8) eingesetzt zur gewünschten Extrapolation führen. Bei dieser Art der Resynthese wird das Prädiktionsfilter anstatt des Prädiktionsfehlerfilters verwendet, d. h. es wird kein Anregungssignal (vgl. Gl. (2.7)) berücksichtigt. Der Grund für die Vernachlässigung der Fehlerenergie liegt darin, dass während der Rekonstruktion keine Werte des Zielsignals vorhanden und somit keine Schätzung des Fehlersignals möglich ist. Deshalb ist die Filterung eines Anregungssignals analog Gl. (2.7) nicht mehr möglich. Die Annahme $e(n) = 0$ wird als eine gängige Alternative unter anderem in diversen Signalrestaurationsalgorithmen vorgeschlagen, siehe beispielsweise [Etter 1996]. Der durch die Überleitung des Prädiktionsfehlerfilters in ein Prädiktionsfilter implizierte Energieverlust führt dazu, dass das extrapolierte Signal gegen Null geht,

$$\hat{x}(n) = \sum_{k=1}^q c(k) x(n-k) \equiv x(n), \quad e(n) = 0. \quad (2.14)$$

Wird der Index der Extrapolation mit $n=l\dots L$ bezeichnet, so führt obige Gleichung zu folgendem System:

$$\begin{pmatrix} x(l) \\ x(l+1) \\ \vdots \\ x(L) \end{pmatrix} = \begin{pmatrix} x(l-1) & x(l-2) & \dots & x(l-q) \\ x(l) & x(l-1) & \dots & x(l-q+1) \\ \vdots & \vdots & \ddots & \vdots \\ x(L-1) & x(L-2) & \dots & x(L-q) \end{pmatrix} \begin{pmatrix} c(1) \\ c(2) \\ \vdots \\ c(q) \end{pmatrix} \quad (2.15)$$

Die Samples $x(l-1)\dots x(l-q)$ entsprechen dabei den q Signalpunkten vor dem Ausfall.

Wird das System mit $x(n) = -\sum_{k=1}^p a(k)x(n-k) + u(n)$ und $u(n) = 0$ als AR-Modell interpretiert, so ist auch hier ersichtlich, dass durch das Fehlen des Anregungssignals $u(n)$ bei stabilen Filterkoeffizienten die Signalenergie gegen Null gehen muss.

Dem Nachteil des Energieverlusts steht der wesentliche Vorteil gegenüber, dass durch diese Art der Filterimplementierung ein stetiger Übergang von Zielsignal zu Extrapolation gewährleistet ist. Wie in der Einleitung erläutert, ist diese Forderung aufgrund der festgelegten Null-Latenz für das zu entwickelnde Verfahren unerlässlich. Bei der Signalrestauration wird der Energieverlust durch die Einführung einer zweiten Extrapolation teilweise kompensiert, bei der mithilfe der dem Ausfall nachfolgenden gültigen Signalpunkten in die entgegengesetzte Richtung (d. h. in die Signalvergangenheit) prädiziert wird. Beide Extrapolationen werden mit einer geeigneten Fensterfunktion multipliziert und können anschließend addiert werden. In Abb. 2-6 ist dies schematisch dargestellt. Die Gln. (2.16) bis (2.18) beschreiben die einzelnen Schritte, wobei die Indizes li bzw. re die links- und rechtsseitige Prädiktion bedeuten.

$$\hat{x}_{li}(n) = \sum_{k=1}^q c_{li}(k)x(n-k), \quad n=l\dots L \quad (2.16)$$

$$\hat{x}_{re}(n) = \sum_{k=1}^q c_{re}(k)x(n+q+1-k), \quad n=L\dots l \quad (2.17)$$

$$\hat{x}(n) = \omega_{li}\hat{x}_{li}(n) + \omega_{re}\hat{x}_{re}(n), \quad n=l\dots L \quad (2.18)$$

Das oberste Bild von Abb. 2-6 zeigt ein mit einem Ausfall behaftetes Signal, darunter ist die Extrapolation dargestellt, bei der die LP-Koeffizienten mit den dem Ausfall vorangehenden Samples geschätzt werden. Die strichliert gezeichnete Amplitudeneinhüllende verdeutlicht die fortschreitende Abnahme der Signalenergie. In Analogie dazu zeigt die dritte Grafik die Extrapolation bezogen auf die der Lücke unmittelbar folgenden Samples, wiederum mit fallender Amplitudencharakteristik. Beide Extrapolationen werden mit einer meist linearen Überblendung zusammengeführt und in der untersten Grafik mit den originalen Signalpunkten verglichen (strichlierte Kurve; die Überblendung ist punktiert eingezeichnet).

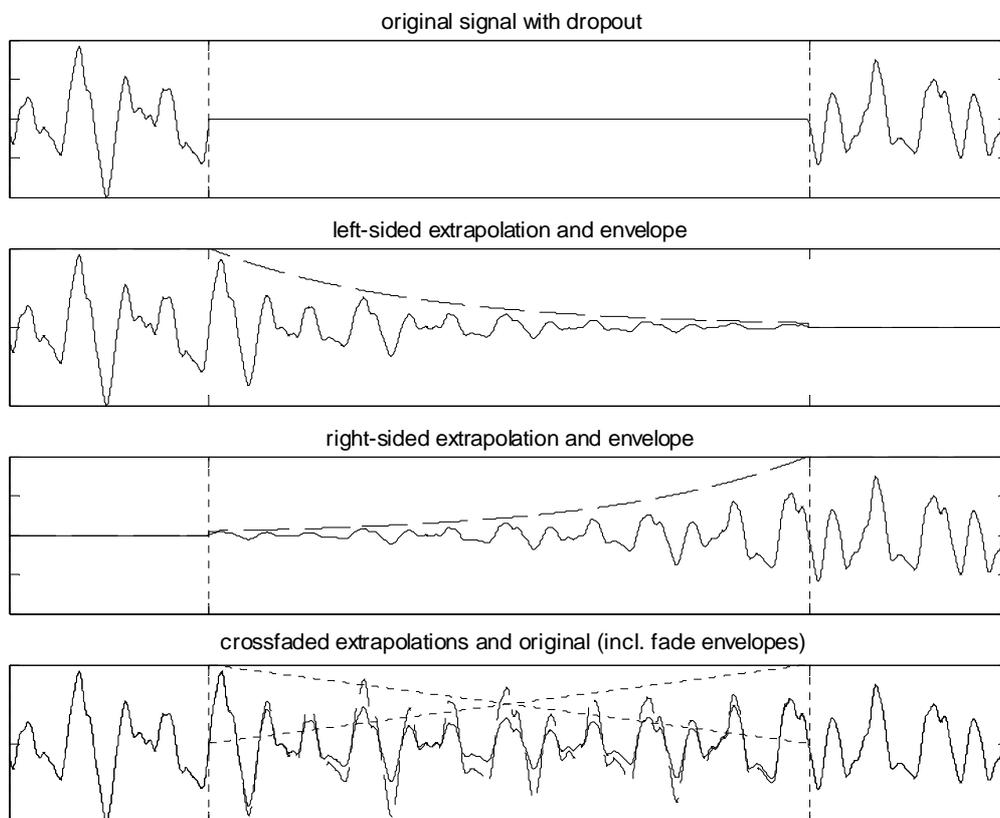


Abb. 2-6: Links- und rechtsseitige Extrapolation

Die Anwendung dieser Technik im E-DROCON-Algorithmus ist nicht möglich, da einerseits die Länge des Ausfalls und das nachfolgende Signal unbekannt sind und andererseits eine Prädiktion in die Vergangenheit bei Echtzeitsystemen unphysikalisch ist.

Eine entsprechende Lösung des Problems wird durch zwei Methoden versucht, die in den nächsten Abschnitten behandelt werden:

- (1) Verbindung der einseitigen Prädiktion mit der Methode der Signalsubstitution: Im Signalpuffer wird nach einem passenden Signalausschnitt gesucht, der parallel zur Prädiktion an die Stelle des Ausfalls gesetzt wird. Mit einer speziellen Überblendungsfunktion werden die beiden Signale überlagert.
- (2) Überführung der Extrapolation in eine Interpolation. Es wird nach einer Möglichkeit gesucht, mit der trotz Echtzeitumgebung eine Art Interpolation realisiert und das Problem des Energieverlusts der Extrapolation kompensiert werden kann.

2.4. Substitution und Interpolation

Der vorliegende Abschnitt untersucht eine Technik zur Rekonstruktion eines fehlenden oder gestörten Signalteils, bei der die unbrauchbaren Samples durch ähnliche aus der Signalvergangenheit ersetzt werden. Dabei sollte der in der Vergangenheit liegende Ausschnitt, der bei Echtzeitrealisationen in einem Puffer gespeichert wird, fehlerfrei sein. In nicht-zeitkritischen Anwendungen stehen im Allgemeinen genügend Abtastpunkte vor und nach dem zu ersetzenden Teil zur Verfügung, d. h. in diesem Fall kann auf beiden Seiten nach dem besten Substitut gesucht werden. Abhängig von der Charakteristik des Signals und der Länge des fehlerhaften Stücks werden zwei Verschleierungsmethoden vorgeschlagen.

- (1) Bei sehr kurzen Aussetzern (wenige Millisekunden) kann der dem Ausfall unmittelbar vorangegangene Signalteil wiederholt werden. Sind Samples zu beiden Seiten des Ausfalls vorhanden, kann eine entsprechende Kombination dieser Samples eingesetzt werden. Bei zu großen Lücken im Signal führt diese Wiederherstellungsmethode jedoch zu hörbaren Artefakten und verschlechtert die Qualität des Ausgangssignals deutlich, [Wasem 1988], [Herre 1995]. Zudem ist sie nicht für den Einsatz in einer Echtzeitumgebung geeignet, da

aufgrund der unbekanntenen Ausfalllänge die zu ersetzenden Signalteile nicht einfach und direkt bestimmt werden können. Wählt man z. B. den ersten Block zu nahe am Ausfallsbeginn, kann es passieren, dass zu wenig nachfolgende Blöcke zur Verfügung stehen. In diesem Fall müsste man entweder dieselben Blöcke nochmals einsetzen oder weiter zurückliegende bestimmen. Beide Kompromisslösungen führen zu unerwünschten Signalverschlechterungen. Andererseits ist es auch nicht sinnvoll, den Abstand zwischen Ausfall und Substitution zu groß zu wählen, weil dadurch die implizit angenommene Quasistationaritätsbedingung ungültig wird und damit ebenfalls kein zufriedenstellendes Ergebnis erreicht wird.

- (2) Ein alternativer Ansatz versucht, diesen Nachteil zu kompensieren, indem eine weitere häufig auftretende Eigenschaft von kurzzeitstationären Signalen ausgenutzt wird. Es wird davon ausgegangen, dass im Signal eine detektierbare Periodizität vorhanden ist und die Periodendauer über eine geeignete Technik ermittelt werden kann. Dementsprechend wird die Lokalisation eines passenden Substituts in die entferntere Vergangenheit verschoben. Zur Ermittlung der Periodizität stehen mehrere Algorithmen zur Verfügung, eine Auswahl davon wurde im E-DROCON-Algorithmus nachgebildet und miteinander verglichen. Die in Kapitel 3 beschriebene dritte subjektive Testreihe dient zur Bestimmung der besten Methode.

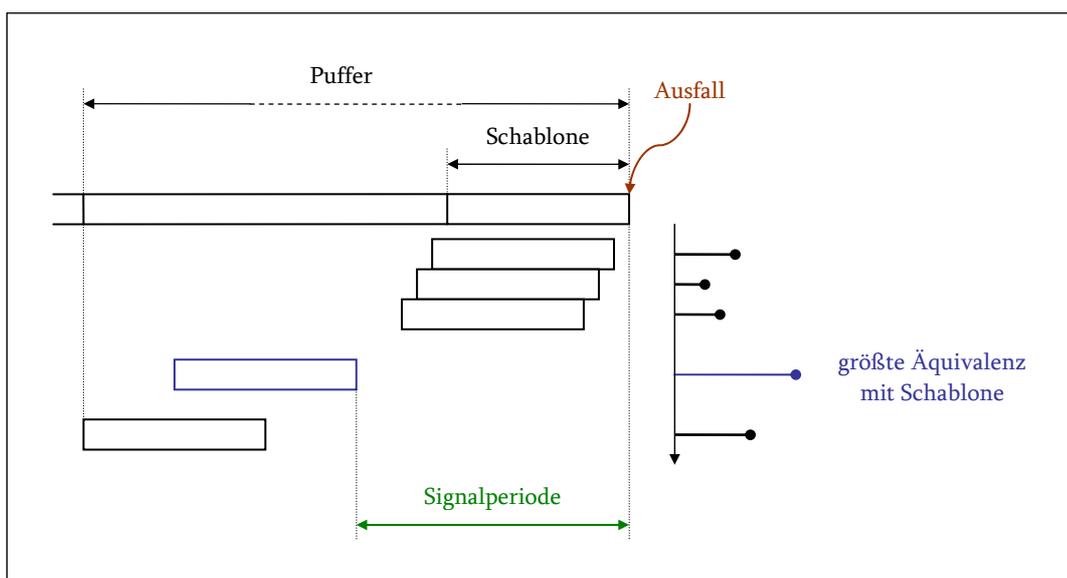


Abb. 2-7: Prinzip der Substitution mit PS

Die Ausnutzung der Signalperiodizität bei der Suche nach den besten Ersatzsamples erweist sich im Hinblick auf die zu verarbeitenden Eingangssignale wie Musik und Sprache als zielführende Entscheidung. Goodman et. al. schlagen in [Goodman 1986] folgende grundlegende Vorgehensweise bei der Wiederherstellung ausgefallener Signalanteile vor:

Der Teil unmittelbar vor dem Ausfall wird als Schablone verwendet und sample- oder blockweise mit dem gesamten Pufferinhalt auf Ähnlichkeit verglichen. Durch die Periodizitätsschätzung wird die Signalperiode geschätzt (siehe Abb. 2-7), die in weiterer Folge dazu dient, einen passenden Ersatz für die ausgefallenen Datenpunkte zu bestimmen.

2.4.1. Periodizitätsschätzung

Die Periodizitätsschätzung (PS) basiert auf den weitverbreiteten Methoden der Ermittlung der Periodendauer T_0 oder Grundfrequenz F_0 bei Sprache, welche Hess in [Hess 1983] als Tonhöhenenerkennung (*pitch determination algorithms*, PDAs) definiert.¹⁰ Diese werden primär klassifiziert in (a) zeitbereichsbasierte Algorithmen und (b) Transformationsverfahren, die auf einer Kurzzeitanalyse des (Zeit-)Signals beruhen. Erstere sind grundsätzlich durch ihre Einfachheit, d. h. geringe Rechenkomplexität gekennzeichnet. Allerdings ist eine zum Teil aufwändige Vorverarbeitung (z. B. Filterung) der Signale notwendig, um eine ausreichend genaue Bestimmung der gewünschten Kenngröße T_0 bzw. F_0 zu erreichen. In der zweiten Klasse geschieht die Analyse des Signals durch eine Transformation in eine Abbildung, die nicht mehr vom Zeitindex abhängig ist. Typischerweise werden dafür Korrelations-, Frequenz- oder Cepstralanalysen eingesetzt.

Nachfolgend werden drei Methoden der PS vorgestellt, welche in den E-DROCON-Algorithmus integriert wurden und zur Realisierung der Signalsubstitution dienen. Eine perzeptive Evaluierung der Qualität von verschleierten Signalen sollte Aufschluss über die Güte der Substitutionstechnik in Verbindung mit jeweils einem der drei Prinzipien der PS

¹⁰ Grundlegendes Ziel der PDAs ist die Ermittlung der Fundamentalfrequenz eines Sprachsignals, die entweder direkt oder über die Periodendauer realisiert werden kann. Im Gegensatz dazu ist für den vorliegenden Anwendungsfall nicht speziell die Periode der tiefsten Frequenzkomponente erforderlich. Dies wäre aufgrund der Komplexität der zulässigen Eingangssignale nicht in dieser Weise realisierbar.

geben. Zur Erzeugung der Testsignale wurden in zunächst fehlerfreie Audiosignale künstlich Ausfälle eingebaut. Anschließend wurden diese mithilfe der drei Varianten der PS bearbeitet, um das Audiosignal wiederherzustellen.

Variante 1: Analyse der Nulldurchgänge¹¹

Das Ziel der Analyse der Nulldurchgänge liegt in der Detektion von strukturellen, sich annähernd regelmäßig wiederholenden Mustern im zeitlichen Signalverlauf. Auf eine Vorfilterung (siehe [Hess 1983]) der Signale wird verzichtet, stattdessen werden Kriterien definiert, die zur Verbesserung des Basisalgorithmus dienen, ohne gleichzeitig den Rechenaufwand zu erhöhen. Im Gegensatz zur ursprünglichen Version, die das Markieren der Nullstellen kontinuierlich ausführt, ist der modifizierte Algorithmus blockorientiert und umfasst folgende Schritte:

(1) Detektion der Nullstellen

Innerhalb der Schablone und in jedem Vergleichsausschnitt des Puffers wird der Zeitindex jedes Nulldurchgangs $\{(n_{1,S}, n_{1,P}^j), (n_{2,S}, n_{2,P}^j), \dots\}$ markiert, die Steigung in diesen Punkten $\{(\Delta_{1,S}, \Delta_{1,P}^j), (\Delta_{2,S}, \Delta_{2,P}^j), \dots\}$ festgehalten und die Gesamtanzahl der Nulldurchgänge (N_S, N_P^j) ermittelt. Die drei Parameter der Vergleichsausschnitte werden mit den entsprechenden der Schablone verglichen. (j ist der Index des Vergleichsausschnitts, S und P kennzeichnen Schablone und Puffer.)

(2) Reihung der Vergleichsausschnitte

Anhand von Kriterien, die unterschiedlich strenge Toleranzbereiche für die in (1) spezifizierten Parameter definieren, werden die Vergleichsausschnitte klassifiziert.

(3) Ermittlung des "besten" Vergleichsausschnitts

Die Gruppe der Vergleichsausschnitte, deren Parameter die geringsten Toleranzen einhalten, wird einer weiteren Prüfung unterzogen, der als Kriterium die Differenz der

¹¹ Aufgrund seiner Einfachheit wurde das Verfahren bereits Mitte der 60er Jahre in analogen Systemen zur Ermittlung der Fundamentalfrequenz bei Sprache eingesetzt. Zufriedenstellende Ergebnisse konnten allerdings nur durch ausreichende Vorfilterung der Signale erzielt werden. Mit der Entwicklung der Digitaltechnik verlor die zeitbereichsbasierte Analyse an Bedeutung und wurde von der Kurzzeitanalyse praktisch abgelöst. Dennoch wird sie nach wie vor als Referenz in Benchmark-Tests herangezogen. Diesen Zweck erfüllt sie auch in der vorliegenden Arbeit.

Signalenergien zugrunde liegt, $\Delta E^k = |E_S - E_P^k|$. Als "bester" Vergleichsausschnitt gilt schließlich jener, dessen ΔE minimal ist. (k ist der Index der Vergleichsausschnitte der in (2) ermittelten "besten" Gruppe.)

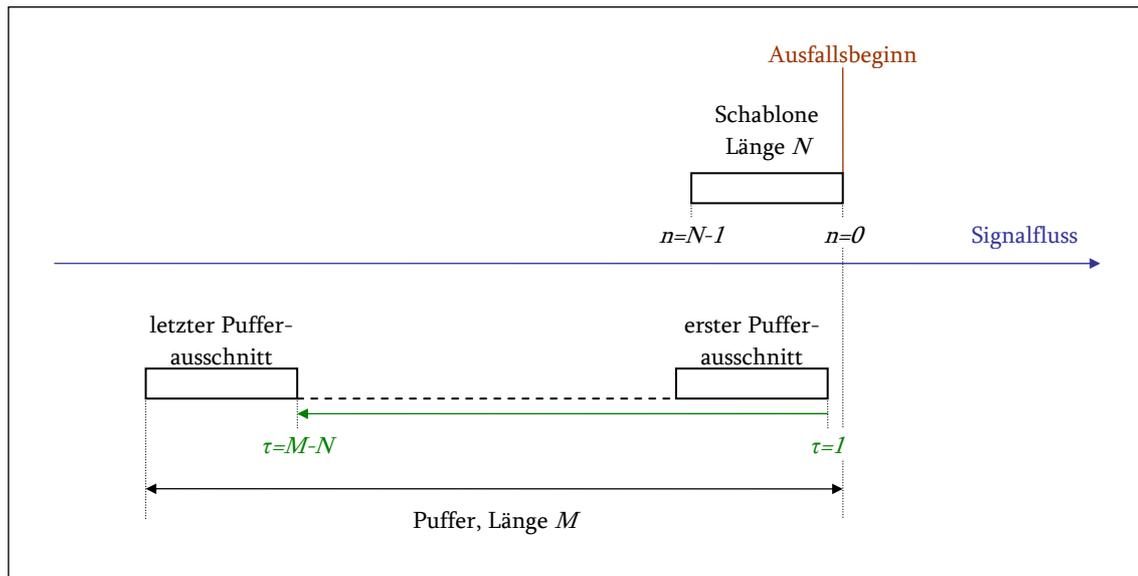
Variante 2: Differenzfunktion nach de Cheveigne und Kawahara

De Cheveigne und Kawahara entwickeln in [de Cheveigne 2002] einen Algorithmus, der von der Autokorrelationsfolge (AKF) ausgeht und durch Modifikationen zur Formulierung einer speziellen Differenzfunktion (DF) führt. Der Ansatz wird als besonders effizient in der Schätzung der Periodizität von komplexen Musiksignalen beschrieben, wenngleich die in der Publikation angeführte Evaluierung ausschließlich mit Sprachsignalen durchgeführt wurde.

Die Korrelation ist ein Maß für die Ähnlichkeit zweier Funktionen und bietet somit eine geeignete Möglichkeit, die Schablone mit dem Inhalt des Puffers zu vergleichen. Da Schablone und Puffer vom selben Signal abgeleitet werden, ist die Betrachtung der AKF zweckmäßig:

$$r(\tau) = \frac{1}{N} \sum_{n=0}^{N-1} x_s(n) x_p(n+\tau). \quad (2.19)$$

Der Index τ repräsentiert zwar eine zeitliche Größe, da er den Versatz an Samples zwischen Schablone x_s und dem jeweiligen Pufferausschnitt x_p darstellt. Dennoch wird die AKF als Transformation in eine auf den Versatz basierte Abbildung, die sogenannte "*lag domain*" interpretiert, in der sie de facto von der tatsächlichen Zeitachse, gegeben durch den Sample-Index n , entkoppelt ist. Diese Unabhängigkeit vereinfacht die Implementierung im E-DROCON-Algorithmus, da der Puffer von der Schablone beginnend sukzessive in die Signalvergangenheit abgearbeitet wird. In Abb. 2-8 ist die Prozedur schematisch dargestellt. Durch die Entkopplung wird das Vorzeichen von τ irrelevant, darum wird dessen Zählrichtung dem Signalfluss entgegengesetzt. Der absolute Wert des Versatzes, bei dem $r(\tau)$ ein Maximum aufweist, wird als Schätzung der Periodizität verwendet.


 Abb. 2-8: Erklärung zur Zählrichtung des Versatzes τ

Weiters bedarf es noch der Klärung eines Details von Gl. (2.19): Die AKF besitzt eine – mit steigendem τ – fallende Amplitude, wenn für ihre Berechnung – wie normalerweise üblich –

– genau ein Signalblock der Länge N zur Verfügung steht: $r'(\tau) = \frac{1}{N} \sum_{n=0}^{N-1-\tau} u(n)u(n+\tau)$. Eine

modifizierte Form der AKF, die Autokovarianzfolge, verzichtet auf eine entsprechende Fensterung und weist dadurch in eine nicht notwendigerweise fallende Amplitude mit

steigendem τ auf: $r''(\tau) = \frac{1}{N} \sum_{n=0}^{N-1} u(n)u(n+\tau)$. Folglich wird im E-DROCON-Algorithmus

streng genommen die Autokovarianzfolge eingesetzt.

Ebenfalls in der *lag domain* definiert ist die gemittelte Amplitudendifferenzfunktion (*average magnitude difference function*, AMDF), die auch die Bezeichnung Antikorrelation trägt, da sie ein Maß für die Unähnlichkeit zweier Funktionen darstellt. Unter der Annahme eines exakt periodischen Signals gilt $\tau = T_0$ und damit weiters folgende Beziehungen:

$$\sum_{n=0}^{N-1} |x_S(n) - x_P(n+T_0)| = 0 \quad \text{bzw.} \quad \sum_{n=0}^{N-1} [x_S(n) - x_P(n+T_0)]^2 = 0.$$

Bei unbekannter Signalperiode lässt sich daraus einerseits die klassische AMDF ableiten:

$$d_1(\tau) = \frac{1}{N} \sum_{n=0}^{N-1} |x_S(n) - x_P(n+\tau)|, \quad (2.20)$$

andererseits kann allgemein ein antagonistischer Zusammenhang zwischen der quadratischen Version der DF und der AKF hergestellt werden:

$$d_2(\tau) = \sum_{n=0}^{N-1} [x_S(n) - x_P(n+\tau)]^2 = r_S(0) + r_P(0, \tau) - 2r(\tau), \quad (2.21)$$

wobei r_S die Signalleistung der Schablone und $r_{P,\tau}$ jene des Pufferausschnitts mit dem Versatz τ zur Schablone bedeuten. Das Vorzeichen des dritten Terms zeigt, dass die beiden Funktionen gegengleich zueinander verlaufen, d. h. große Werte von $r(\tau)$ entsprechen kleinen Werten von $d_2(\tau)$, allerdings müssen die beiden Extremwerte nicht exakt an der selben Stelle τ auftreten. Der Grund für eine Abweichung ist die Abhängigkeit des zweiten Terms ebenfalls von τ . Diese Eigenschaft führt zu einer deutlich verbesserten PS bei Verwendung von Gl. (2.21) anstatt Gl. (2.19). De Cheveigne und Kawahara schreiben von einer Reduktion der Fehlerrate von bis zu 1/5 im Vergleich von AKF und DF. Darüber hinaus zeigt sich, dass die DF unempfindlicher auf Amplitudenschwankungen der nichtstationären Eingangssignale reagiert als die AKF. Dieser Vorteil kann durch eine weitere Modifikation verbessert werden und führt zur kumulativen mittleren normalisierten Differenzfunktion (*cumulative mean normalized difference function*, CMNDF):

$$d_3(\tau) = \frac{d_2(\tau)}{\frac{1}{\tau} \sum_{j=1}^{\tau} d_2(j)} \quad \tau > 0. \quad (2.22)$$

An der Stelle $\tau = 0$ wird die CMNDF definitionsgemäß auf Eins gesetzt: $d_3(0) = 1$.

Variante 3: Differenzfunktion nach Goodman et. al.

Goodman et. al. vergleichen in [Goodman 1986] mehrere Methoden des Mustervergleichs (Kurzzeitanalyse) und der Tonhöhenenerkennung (Analyse im Zeitbereich). Erstere umfassen die normierte (modifizierte) Autokorrelationsfolge sowie die Differenzfunktion mit unterschiedlichen Normierungen. Auf letztere wird nicht weiter eingegangen, da sie einen ähnlichen Algorithmus wie Variante 1 beschreibt, allerdings die Detektion von Spitzenwerten (Maxima und Minima der Signalamplitude) anstatt Nullstellen vornimmt.

Die leistungsbezogene Normierung der AKF lautet:

$$r_N(\tau) = \frac{\sum_{n=0}^{N-1} x_S(n) x_P(n+\tau)}{\sqrt{\sum_{n=0}^{N-1} x_S(n)^2 x_P(n+\tau)^2}}. \quad (2.23)$$

Als Alternative werden drei DFn vorgeschlagen:

$$d_4(\tau) = \frac{1}{N} \sum_{n=0}^{N-1} \left| \frac{x_S(n)}{\sqrt{\sum_{j=0}^{N-1} x_S(j)^2}} - \frac{x_P(n+\tau)}{\sqrt{\sum_{j=0}^{N-1} x_P(j+\tau)^2}} \right| \quad (2.24)$$

$$d_5(\tau) = \frac{1}{N} \sum_{n=0}^{N-1} \left| \frac{x_S(n)}{\sum_{j=0}^{N-1} |x_S(j)|} - \frac{x_P(n+\tau)}{\sum_{j=0}^{N-1} |x_P(j+\tau)|} \right| \quad (2.25)$$

$$d_6(\tau) = \frac{1}{N} \sum_{n=0}^{N-1} \left| \frac{x_S(n)}{x_{S,\max} - x_{S,\min}} - \frac{x_P(n+\tau)}{x_{P,\max} - x_{P,\min}} \right| \quad (2.26)$$

Ein informeller Vergleich wurde zur Vorselektion einer der vier Formulierungen Gl. (2.23) bis Gl. (2.26) angestellt und zeigte, dass Gl. (2.24) bzw. Gl. (2.25) ähnlich gute Ergebnisse lieferten, während die Fehlerrate bei Gl. (2.23) und Gl. (2.26) höher war. Aufgrund des geringeren Rechenaufwandes wurde Gl. (2.25) vorgezogen und als beste der vier Berechnungsmethoden festgelegt.

Somit wird die perzeptive Evaluierung mit

- V1: Nullstellenanalyse,
- V2: CMNDF nach Gl. (2.22) und
- V3: DF nach Gl. (2.25) durchgeführt.

2.4.2. Substitution

Mit der Schätzung der Periodizität wird letztendlich das Substitut gebildet, welches bei *offline*, d. h. nicht echtzeitfähigen Anwendungen in seiner Länge genau auf den zu ersetzenden

Signalteil abgestimmt werden kann. Im vorliegenden Fall wird ab der Detektion eines Ausfalls ein Block nach dem anderen vom Puffer ausgelesen und an das aktuelle Ende des Signals angehängt. Dieser Vorgang wiederholt sich bis zur Verfügbarkeit des nächsten nicht ausgefallenen Blocks des Zielsignals. Es wird auch bei dieser Technik ersichtlich, dass die beiden kritischen Übergangsstellen unstetige Signalverläufe produzieren.

Der Wechsel von der Substitution zum wiedervorhandenen Zielsignal kann durch eine Überblendung realisiert werden, indem während der ersten gültigen Blöcke des Zielsignals die Auslesung aus dem Puffer fortgeführt wird, damit ein entsprechendes Überblenden einen kontinuierlichen Signalfluss gewährleistet. Speziell am Beginn des Ausfalls gibt es jedoch aufgrund des sofortigen Signalverlustes keine Möglichkeit für eine Überblendung.

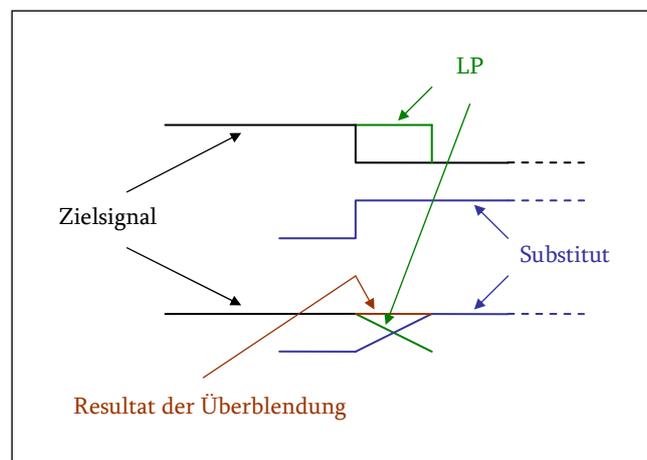


Abb. 2-9: Stetiger Übergang vom Zielsignal zum Substitut

Im vorigen Abschnitt wurde gezeigt, dass die lineare Prädiktion zwar den großen Vorteil einer stetigen Extrapolation besitzt, durch den der Berechnung inhärenten Energieverlust jedoch nicht für eine längere Ausfallsverschleierung geeignet ist. Es stellt sich die Frage, ob die Vorteile beider Methoden zusammengefasst und gemeinsam verwertet werden können. Abb. 2-9 veranschaulicht diese Idee. Mithilfe der LP wird das Zielsignal um einen Bereich verlängert, der für ein zufriedenstellendes Überblenden zur Substitution erforderlich ist. Diese beginnt gleichzeitig mit dem ersten LP-Block, wobei beide Signale mit einer passenden Fensterfunktion multipliziert und anschließend addiert werden. Die Realisierung der Überblendung erlaubt einen kontinuierlichen Übergang vom extrapolierten Zielsignal auf

dessen Schätzung, das aus der Vergangenheit, d. h. dem Puffer, genommen wird, und ist in dieser Form in ein Echtzeitsystem integrierbar.

2.4.3. Interpolation mittels AR-Modell

Eine Zusammenfassung der beschriebenen Prinzipien ergibt, dass lediglich die Kombination aus Extrapolation und Substitution zu einer optimierten Lösung für eine Ausfallsverschleierung führt, die die vorliegenden Anforderungen erfüllt. In diesem Abschnitt wird gezeigt, wie beide Techniken in einen Algorithmus integriert werden können. Dazu ist es notwendig, das lineare Vorwärtsprädiktionsfehlerfilter nochmals genau zu betrachten. Die Berechnung des Fehlers nach Gl. (2.9) wird nun unter der Annahme einer blockweisen Verarbeitung der Signalfolge durchgeführt, wobei die Blocklänge mit N und die Prädiktionsordnung mit q angegeben wird [Vaseghi 1996].¹²

$$\begin{pmatrix} e(1) \\ \vdots \\ e(q+1) \\ \vdots \\ e(N-q) \\ \vdots \\ e(N) \\ \vdots \\ e(N+q) \end{pmatrix} = \begin{pmatrix} x(1) & \dots & 0 \\ \vdots & \ddots & \vdots \\ x(q+1) & \dots & x(1) \\ \vdots & \ddots & \vdots \\ x(N-q) & \dots & x(q+1) \\ \vdots & \ddots & \vdots \\ x(N) & \dots & x(N-q) \\ \vdots & \ddots & \vdots \\ 0 & \dots & x(N) \end{pmatrix} \begin{pmatrix} 1 \\ -w(1) \\ \vdots \\ -w(q) \end{pmatrix} \quad (2.27)$$

Die Signalmatrix repräsentiert eine $(N+q) \times (q+1)$ Toeplitz-Matrix und kann in drei Komponenten zerlegt werden:

$$\mathbf{X} = \begin{pmatrix} \mathbf{X}_{pre} \\ \mathbf{X}_{full} \\ \mathbf{X}_{post} \end{pmatrix} \quad (2.28)$$

¹² Die in diesem Abschnitt definierten Berechnungen beziehen sich ausschließlich auf Systeme zur Prädiktion von in der Zukunft liegenden Samples (Vorwärtsprädiktion). Deshalb wird hier auf den Index f der Filterkoeffizienten verzichtet.

wobei \mathbf{X}_{pre} und \mathbf{X}_{post} jeweils eine untere bzw. obere Dreiecksmatrix der Dimension $q \times (q+1)$ ergeben; \mathbf{X}_{full} ist eine $(N-q) \times (q+1)$ Rechtecksmatrix. Die Indizes *pre* und *post* beziehen sich auf die angefügten Nullen zu Beginn bzw. am Ende des Signalvektors, *full* weist auf eine vollständige Signalmatrix (ohne Nullen) hin.

$$\begin{aligned}
 \mathbf{X}_{pre} &= \begin{pmatrix} x(1) & \dots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ x(q) & \dots & x(1) & 0 \end{pmatrix} \\
 \mathbf{X}_{full} &= \begin{pmatrix} x(q+1) & \dots & x(1) \\ \vdots & \ddots & \vdots \\ x(N-q) & \dots & x(q+1) \\ \vdots & \ddots & \vdots \\ x(N) & \dots & x(N-q) \end{pmatrix} \\
 \mathbf{X}_{post} &= \begin{pmatrix} 0 & x(N) & \dots & x(N-q+1) \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & x(N) \end{pmatrix}
 \end{aligned} \tag{2.29}$$

Aus den beiden Dreiecksmatrizen ist ersichtlich, dass jene Elemente von \mathbf{X} mit Nullen aufgefüllt werden, die aufgrund der Definition des Signalvektors $[x(1) \dots x(N)]^T$ nicht verfügbar sind.

Mit der kompakten Notation

$$\mathbf{e}_x = \mathbf{X} \begin{pmatrix} 1 \\ -\mathbf{w} \end{pmatrix} \tag{2.30}$$

erhält man den quadratischen Fehler

$$P = E \left\{ \sum |e(n)|^2 \right\} = E \{ \mathbf{e}^H \mathbf{e} \}. \tag{2.31}$$

Je nach Wahl der Summationsgrenzen in Gl. (2.31) unterscheidet man folgende vier Fälle:

- *windowed* (Autokorrelationsmethode),
- *non-windowed* (Kovarianzmethode),
- *pre-windowed*,

– *post-windowed*.

Für die Berechnung des Quadrats des gesamten Vektors $\mathbf{e}_x = [e(1) \ \dots \ e(N+q)]^T$ bei der oben genannten Blockverarbeitung wird die Signalfolge vor und nach den gültigen Datensamples mit Nullen aufgefüllt. Dieser Fall wird als Autokorrelationsmethode der linearen Prädiktion bezeichnet. Wird nur die reduzierte Signalmatrix \mathbf{X}_{full} verwendet, verkürzt sich dementsprechend der Fehlervektor zu $\mathbf{e}_{full} = [e(q+1) \ \dots \ e(N)]^T$. Man spricht in diesem Fall von der ungefensterten oder der Kovarianzmethode der linearen Prädiktion. Die Bezeichnungen "Autokorrelationsmethode" und "Kovarianzmethode" stammen aus der Sprachverarbeitung. Weitere Details zur Sprachverarbeitung finden sich u. a. in [Deller 2000], in dieser Arbeit wird nicht näher darauf eingegangen. Als dritte Variante wird jener Fall berücksichtigt, bei dem der Fehlervektor exakt auf die Blocklänge abgestimmt ist, $\mathbf{e}_{pre} = [e(1) \ \dots \ e(N)]^T$. Diese *pre-windowed* Methode impliziert eine Fensterung am Beginn des Datenvektors, weshalb das resultierende System die Matrizen \mathbf{X}_{pre} und \mathbf{X}_{full} beinhaltet. Umgekehrt entspricht die Verwendung der Matrizen \mathbf{X}_{full} und \mathbf{X}_{post} einer Fensterung am Ende des Signalblocks (*post-windowed*) und der Fehlervektor umfasst in diesem Fall die Samples $\mathbf{e}_{post} = [e(q+1) \ \dots \ e(N+q)]^T$. Je nach Verfügbarkeit der Signalsamples und Zieldefinition der Berechnung liefern die einzelnen Techniken mehr oder weniger optimale Ergebnisse.

Die hier behandelte Problemstellung kann auf verschiedene Arten interpretiert werden, einerseits unterscheidet man als fundamentale Methoden der Signalfortführung zwischen Extrapolation und Interpolation. Der zugrunde liegende Algorithmus ist in beiden Fällen zweistufig aufgebaut, wobei die erste Stufe die Schätzung der Filterkoeffizienten vornimmt und die zweite für die Erzeugung der fehlenden Signalsamples verantwortlich ist. Eine weitere Einteilung differenziert Systeme, in denen (a) beide Stufen gemeinsam betrachtet und (b) diese in getrennten Schritten ausgeführt werden. Abb. 2-10 zeigt die sich daraus ergebenden vier Möglichkeiten und deren Realisierungsmethoden.

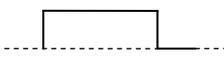
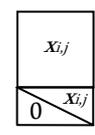
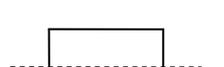
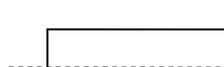
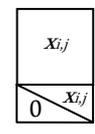
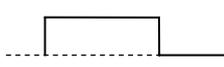
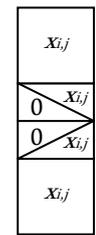
	Extrapolation	Interpolation
(a) Stufen gemeinsam betrachtet	<i>post-windowed</i>  	Kovarianz (incl. Ausfall/Nullen)  
(b) Stufen getrennt betrachtet	Kovarianz  	Kovarianz  
Koeffizientenschätzung		
Signalerzeugung	<i>post-windowed</i>  	<i>post- & pre-windowed</i>  

Abb. 2-10: Einteilung der Prädiktionssysteme

In der Grafik ist die Signalfolge sowie die Struktur der jeweiligen Signalmatrix skizziert. Der Vergleich der beiden Varianten der Extrapolation lässt beispielsweise erkennen, dass im Fall (a) die notwendige Nullfolge am Signalende auch die Schätzung der Koeffizienten beeinflusst, während dies beim getrennten System (b) vermieden werden kann. Laut Gl. (2.30), hängen Schätzung der Filterkoeffizienten und Fehlervektor über die Signalmatrix zusammen, wobei bei der Kovarianzmethode ausschließlich tatsächliche Signalsamples in Betracht gezogen werden:

$$\mathbf{e}_{full} = \mathbf{X}_{full} \begin{pmatrix} 1 \\ -\mathbf{w} \end{pmatrix}. \quad (2.32)$$

Bei der *post-windowed*-Methode hingegen werden dem Signal Nullen angefügt. Dies induziert einen zusätzlichen Approximationsfehler, der sich negativ auf die Ermittlung der Koeffizienten auswirkt,

$$\mathbf{e}_{post} = \begin{pmatrix} \mathbf{X}_{full} \\ \mathbf{X}_{post} \end{pmatrix} \begin{pmatrix} 1 \\ -\mathbf{w} \end{pmatrix}. \quad (2.33)$$

Allgemein erweist sich daher die Kovarianzmethode als gegenüber der *post-windowed*-Methode bessere Realisierung. Dennoch findet der *post-windowed*-Fall Anwendung in Extrapolationsalgorithmen [Vaseghi 1996], da eine deutliche Äquivalenz zu deren Problemstellung besteht, siehe Abb. 2-11. Die Verlängerung einer Folge mit Nullen (linke Grafik) wird dem Signalausfall gleichgesetzt (rechte Grafik), womit sich rechts eine neue Blocklänge $N' = N - L$ definieren lässt. Die Anzahl der Nullen ist durch die Ausfallslänge L bestimmt und ersetzt in Folge den Parameter q . [Etter 1996], [Janssen 1986] oder [Veldhuis 1988] verwenden diese Art der Implementierung und betrachten den Fehlervektor aus Gl. (2.33) als eine von zwei Variablen abhängige Größe $\mathbf{e}_{post} = Fkt\{\mathbf{w}, \mathbf{x}_u\}$. Damit definieren sie ein Gesamtsystem analog der Interpolation, Fall (a) (Abb. 2-10). In einem zweistufigen Verfahren werden zuerst die Filterkoeffizienten \mathbf{w} und anschließend die unbekanntes Samples \mathbf{x}_u des zugrunde liegenden Gleichungssystems ermittelt. Der Vorteil gegenüber der Extrapolationsmethode aus Abschnitt 2.3.3, (entspricht Fall (b)) liegt darin, dass Struktur und Dimension des Systems in beiden Verarbeitungsschritten unverändert bleiben.

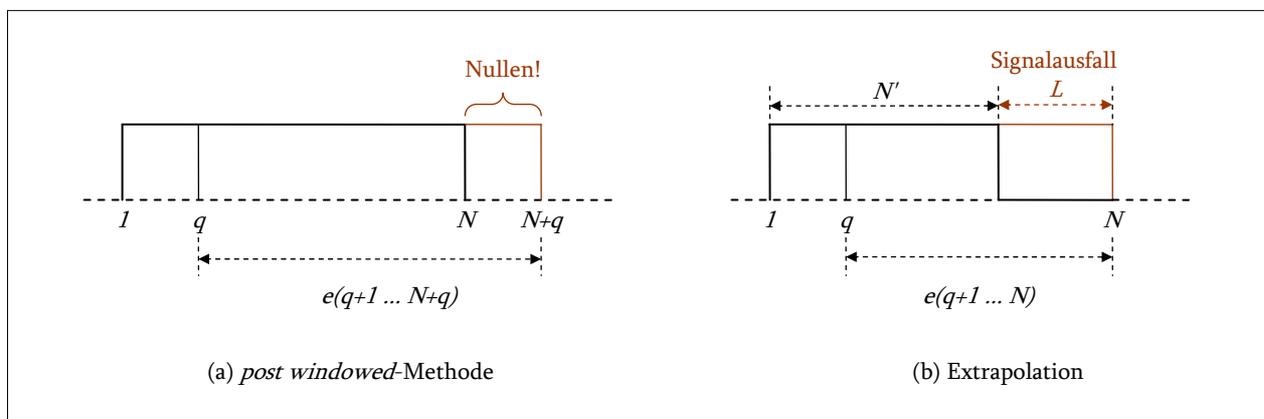


Abb. 2-11: Vergleich von *post-windowed*-Methode und Extrapolation

Die Interpolation¹³ kann ebenfalls auf beide Arten ausgeführt werden. In der vorliegenden Arbeit wird das zweiteilige System als alternative Verschleierungsmethode zur Kombination aus Extrapolation und Substitution realisiert. Bevor speziell auf diese eingegangen wird, erfolgt eine Diskussion über die allgemeinen Eigenschaften der Interpolation unter der

¹³ Vaseghi verwendet in [Vaseghi 1996] die Bezeichnung "statistische Interpolation" für jene Arten der Interpolation, die sich auf die Auswertung statistischer Parameter von (i. A.) Zufallsfolgen stützt.

Voraussetzung, dass zu beiden Seiten der Lücke tatsächlich genügend Signalsamples für die Berechnung zur Verfügung stehen.

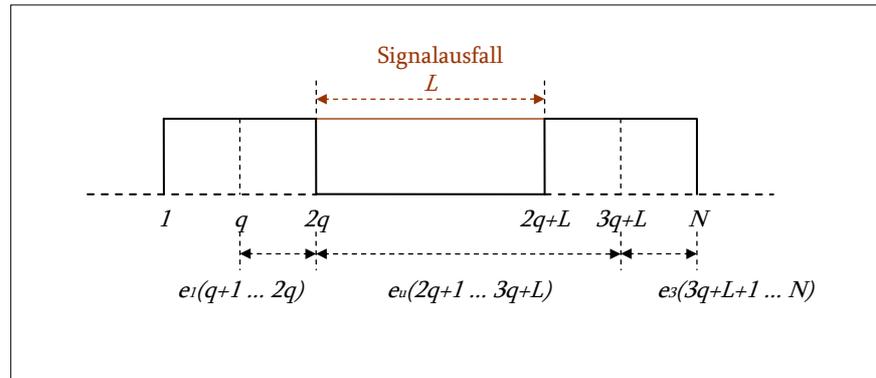


Abb. 2-12: Interpolation, Zusammensetzung eines Signalblocks

In Abb. 2-12 ist die Zusammensetzung eines Signalblocks bei der Interpolation skizziert, wobei wiederum die Gesamtlänge mit N bezeichnet wird. Es ist ersichtlich, dass der Signalausfall in der Mitte des Blocks liegt und zu beiden Seiten $2q$ Samples des Zielsignals vorhanden sind. Aus dieser Tatsache heraus kann das Gleichungssystem aus Gl. (2.32) folgendermaßen unterteilt werden:

$$\begin{pmatrix} e(q+1) \\ \vdots \\ e(2q) \\ \hline e(2q+1) \\ \vdots \\ e(3q+L) \\ \hline e(3q+L+1) \\ \vdots \\ e(N) \end{pmatrix} = \begin{pmatrix} x(q+1) & \dots & x(1) \\ \vdots & \ddots & \vdots \\ x(2q) & \dots & x(q) \\ \hline x_u(2q+1) & \dots & x(q+1) \\ \vdots & \ddots & \vdots \\ x(3q+L) & \dots & x_u(2q+L) \\ \hline x(3q+L+1) & \dots & x(2q+L+1) \\ \vdots & \ddots & \vdots \\ x(N) & \dots & x(N-q) \end{pmatrix} \begin{pmatrix} 1 \\ -w(1) \\ \vdots \\ -w(q) \end{pmatrix} \quad (2.34)$$

Erster und letzter Teil von Gl. (2.34) beinhalten ausschließlich verfügbare Signalsamples, während der mittlere Teil des Systems mit unbekanntem $x_u(n)$ (bzw. Nullsamples) behaftet ist. Um eine Minimierung des quadratischen Fehlers P aus Gl. (2.31) zu erreichen, wird zunächst Gl. (2.32) in der erweiterten Form

$$\mathbf{e}_{full} = \mathbf{X}_{full} \begin{pmatrix} 1 \\ -\mathbf{w} \end{pmatrix} = \begin{pmatrix} \mathbf{x} & \tilde{\mathbf{X}}_{full} \end{pmatrix} \begin{pmatrix} 1 \\ -\mathbf{w} \end{pmatrix} = \mathbf{x} - \tilde{\mathbf{X}}_{full} \mathbf{w} \quad (2.35)$$

verwendet. Unter Verwendung von Gl. (2.35) wird der quadratische Fehler in Gl. (2.31) zu

$$E\{\mathbf{e}^H \mathbf{e}\} = E\{\mathbf{x}^H \mathbf{x}\} - 2\mathbf{w}^H E\{\tilde{\mathbf{X}}_{full}^H \mathbf{x}\} + \mathbf{w}^H E\{\tilde{\mathbf{X}}_{full}^H \tilde{\mathbf{X}}_{full}\} \mathbf{w}. \quad (2.36)$$

Darin besteht der Vektor \mathbf{x} aus den Elementen $\mathbf{x} = [x(q+1) \dots x(N)]^T$, die Matrix $\tilde{\mathbf{X}}_{full}$ repräsentiert die um \mathbf{x} reduzierte Matrix \mathbf{X}_{full} . Da der Fehlervektor eine von den zwei Variablen \mathbf{x}_u und \mathbf{w} abhängige Funktion und \mathbf{x}_u aufgrund obiger Definitionen in \mathbf{x} und $\tilde{\mathbf{X}}_{full}$ enthalten ist, muss Gl. (2.36) nach \mathbf{w} , \mathbf{x} und $\tilde{\mathbf{X}}_{full}$ abgeleitet werden. Die Lösung dieses mathematischen Problems ist nur mit einem nichtlinearen Ansatz durchführbar. Anstelle aufwendiger Berechnungsverfahren verspricht folgende Prozedur eine ausreichend genaue Näherung [Veldhuis 1988], [Vaseghi 1996]. Zunächst wird angenommen, dass die unbekanntes Samples durch Nullen ersetzt werden. Dies entspricht Fall (a) bei der Interpolation, siehe Abb. 2-10 rechts oben. Somit kann Gl. (2.36) ausschließlich nach \mathbf{w} abgeleitet werden:

$$-2E\{\tilde{\mathbf{X}}_{full}^H \mathbf{x}\} + 2E\{\tilde{\mathbf{X}}_{full}^H \tilde{\mathbf{X}}_{full}\} \mathbf{w} \doteq \mathbf{0}. \quad (2.37)$$

Die Filterkoeffizienten werden in Folge aus den mit Nullen aufgefüllten bekannten Signalsamples geschätzt:

$$\mathbf{w} = E\{\tilde{\mathbf{X}}_{full}^H \tilde{\mathbf{X}}_{full}\}^{-1} E\{\tilde{\mathbf{X}}_{full}^H \mathbf{x}\}. \quad (2.38)$$

Die Annahme $\mathbf{x}_u = 0$ bringt eine Verschlechterung der Approximation mit sich, die vorerst toleriert wird. Unter Verwendung der Autokorrelationsmatrix $\tilde{\mathbf{\Phi}} = E\{\tilde{\mathbf{X}}_{full}^H \tilde{\mathbf{X}}_{full}\}$ und des Korrelationsvektors $\tilde{\mathbf{\phi}} = E\{\tilde{\mathbf{X}}_{full}^H \mathbf{x}\}$ lautet die Lösung in kompakter Form:

$$\mathbf{w} = \tilde{\mathbf{\Phi}}^{-1} \tilde{\mathbf{\phi}}. \quad (2.39)$$

In der zweiten Stufe des Algorithmus wird die Interpolation realisiert, wozu Gl. (2.32) bzw. Gl. (2.34) neu arrangiert wird:

$$\mathbf{e}_{full} = \mathbf{W}_u \mathbf{x}_u + \mathbf{W}_1 \mathbf{x}_1 + \mathbf{W}_3 \mathbf{x}_3. \quad (2.40)$$

Der Aufbau der drei Teilsysteme wird am besten anhand eines konkreten Beispiels mit $q = 3$ und $L = 4$ verdeutlicht (als Bezug wird auf der linken Seite jeweils auch der Fehlervektor angeführt):

$$\mathbf{W}_u \mathbf{x}_u = \begin{pmatrix} e(q+1) \\ e(q+2) \\ e(2q) \\ \hline e(2q+1) \\ e(2q+2) \\ e(2q+3) \\ e(2q+L) \\ \hline e(2q+L+1) \\ e(2q+L+2) \\ e(3q+L) \\ \hline e(3q+L+1) \\ e(3q+L+2) \\ e(N) \end{pmatrix} \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \hline 1 & 0 & 0 & 0 \\ -w(1) & 1 & 0 & 0 \\ -w(2) & -w(1) & 1 & 0 \\ -w(q) & -w(2) & -w(1) & 1 \\ \hline 0 & -w(q) & -w(2) & -w(1) \\ 0 & 0 & -w(q) & -w(2) \\ 0 & 0 & 0 & -w(q) \\ \hline 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} x_u(q+1) \\ x_u(q+2) \\ x_u(q+3) \\ x_u(q+L) \end{pmatrix}$$

$$\mathbf{W}_1 \mathbf{x}_1 = \begin{pmatrix} e(q+1) \\ e(q+2) \\ e(2q) \\ \hline e(2q+1) \\ e(2q+2) \\ e(2q+3) \\ e(2q+L) \\ \hline e(2q+L+1) \\ e(2q+L+2) \\ e(3q+L) \\ \hline e(3q+L+1) \\ e(3q+L+2) \\ e(N) \end{pmatrix} \begin{pmatrix} -w(q) & -w(2) & -w(1) & 1 & 0 & 0 \\ 0 & -w(q) & -w(2) & -w(1) & 1 & 0 \\ 0 & 0 & -w(q) & -w(2) & -w(1) & 1 \\ \hline 0 & 0 & 0 & -w(q) & -w(2) & -w(1) \\ 0 & 0 & 0 & 0 & -w(q) & -w(2) \\ 0 & 0 & 0 & 0 & 0 & -w(q) \\ 0 & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} x(1) \\ x(2) \\ x(q) \\ x(q+1) \\ x(q+2) \\ x(2q) \end{pmatrix}$$

$$\begin{pmatrix} e(q+1) \\ e(q+2) \\ e(2q) \\ \hline e(2q+1) \\ e(2q+2) \\ e(2q+3) \\ e(2q+L) \\ \hline e(2q+L+1) \\ e(2q+L+2) \\ e(3q+L) \\ \hline e(3q+L+1) \\ e(3q+L+2) \\ e(N) \end{pmatrix} \mathbf{W}_3 \mathbf{x}_3 = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ \hline 1 & 0 & 0 & 0 & 0 & 0 \\ -w(1) & 1 & 0 & 0 & 0 & 0 \\ -w(2) & -w(1) & 1 & 0 & 0 & 0 \\ \hline -w(q) & -w(2) & -w(1) & 1 & 0 & 0 \\ 0 & -w(q) & -w(2) & -w(1) & 1 & 0 \\ 0 & 0 & -w(q) & -w(2) & -w(1) & 1 \end{pmatrix} \begin{pmatrix} x(2q+L+1) \\ x(2q+L+2) \\ x(3q+L) \\ x(3q+L+1) \\ x(3q+L+2) \\ x(N) \end{pmatrix}$$

Die Ermittlung des unbekanntem Vektors erfolgt wiederum über die Minimierung des quadratischen Fehlers P , der nun mittels Gl. (2.40) gebildet, nach \mathbf{x}_u abgeleitet und Null gesetzt wird:

$$P = \mathbf{e}_{full}^H \mathbf{e}_{full} = (\mathbf{W}_u \mathbf{x}_u + \mathbf{W}_1 \mathbf{x}_1 + \mathbf{W}_3 \mathbf{x}_3)^H (\mathbf{W}_u \mathbf{x}_u + \mathbf{W}_1 \mathbf{x}_1 + \mathbf{W}_3 \mathbf{x}_3), \quad (2.41)$$

Als Endergebnis der statistischen Interpolationsmethode erhält man eine *least squares*-Schätzung der fehlenden Signalpunkte innerhalb der beobachteten Folge $[x(1) \dots x(N)]^T$:

$$\hat{\mathbf{x}}_{u,a} = -(\mathbf{W}_u^H \mathbf{W}_u)^{-1} [\mathbf{W}_u^H (\mathbf{W}_1 \mathbf{x}_1 + \mathbf{W}_3 \mathbf{x}_3)]. \quad (2.42)$$

Eine Verbesserung des Ergebnisses wird erzielt, indem man eine Aufteilung des Fehlervektors \mathbf{e}_{full} in $\mathbf{e}_{full} = [\mathbf{e}_1 \quad \mathbf{e}_u \quad \mathbf{e}_3]^T$ vornimmt und Gl. (2.38) dahingehend modifiziert, dass die Gewinnung von \mathbf{w} ausschließlich auf dem ersten und dritten Teilsystem aus Gl. (2.34) basiert. Infolgedessen sind vor und nach dem Ausfall mindestens $2q$ gültige Samples notwendig, mit denen die Berechnung der Koeffizienten analog Gl. (2.38) ausgeführt wird, d.

$$\text{h.} \quad \mathbf{x} \rightarrow [\mathbf{x}_1 \mid \mathbf{x}_3]^T, \quad \tilde{\mathbf{X}}_{full} \rightarrow [\tilde{\mathbf{X}}_1 \mid \tilde{\mathbf{X}}_3]^T, \quad \text{mit} \quad \mathbf{x}_1 = [x(q+1) \ \dots \ x(2q)]^T,$$

$$\mathbf{x}_3 = [x(3q+L+1) \ \dots \ x(N)]^T, \quad \text{sowie} \quad \text{den} \quad \text{quadratischen} \quad \text{Toeplitz-Matrizen}$$

$$\tilde{\mathbf{X}}_1 = \begin{pmatrix} x(q) & \dots & x(1) \\ \vdots & \ddots & \vdots \\ x(2q-1) & \dots & x(q) \end{pmatrix} \quad \text{und} \quad \tilde{\mathbf{X}}_3 = \begin{pmatrix} x(3q+L) & \dots & x(2q+L+1) \\ \vdots & \ddots & \vdots \\ x(N-1) & \dots & x(3q+L) \end{pmatrix}. \quad \text{Mit} \quad \text{diesen}$$

reduzierten Systemen erhält man die Filterkoeffizienten

$$\tilde{\mathbf{w}} = \tilde{\Phi}^{-1} \tilde{\phi} = \left[\begin{pmatrix} \tilde{\mathbf{X}}_1 \\ \tilde{\mathbf{X}}_3 \end{pmatrix}^H \begin{pmatrix} \tilde{\mathbf{X}}_1 \\ \tilde{\mathbf{X}}_3 \end{pmatrix} \right]^{-1} \begin{pmatrix} \tilde{\mathbf{X}}_1 \\ \tilde{\mathbf{X}}_3 \end{pmatrix}^H \begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_3 \end{pmatrix}. \quad (2.43)$$

Das System zur Ermittlung des fehlenden Signalteils wird nun exakt auf den mittleren Teil aus Gl. (2.34) zugeschnitten, die drei Teilsysteme vereinfachen sich dementsprechend zu (wiederum bezogen auf das Beispiel mit $q = 3$ und $L = 4$):

$$\tilde{\mathbf{W}}_u \mathbf{x}_u = \begin{pmatrix} e(2q+1) \\ e(2q+2) \\ e(2q+3) \\ e(2q+L) \\ \hline e(2q+L+1) \\ e(2q+L+2) \\ e(3q+L) \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ -w(1) & 1 & 0 & 0 \\ -w(2) & -w(1) & 1 & 0 \\ -w(q) & -w(2) & -w(1) & 1 \\ \hline 0 & -w(q) & -w(2) & -w(1) \\ 0 & 0 & -w(q) & -w(2) \\ 0 & 0 & 0 & -w(q) \end{pmatrix} \begin{pmatrix} x_u(q+1) \\ x_u(q+2) \\ x_u(q+3) \\ x_u(q+L) \end{pmatrix},$$

$$\tilde{\mathbf{W}}_1 \mathbf{x}_1 = \begin{pmatrix} e(2q+1) \\ e(2q+2) \\ e(2q+3) \\ e(2q+L) \\ \hline e(2q+L+1) \\ e(2q+L+2) \\ e(3q+L) \end{pmatrix} = \begin{pmatrix} -w(q) & -w(2) & -w(1) \\ 0 & -w(q) & -w(2) \\ 0 & 0 & -w(q) \\ \hline 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} x(q+1) \\ x(q+2) \\ x(2q) \end{pmatrix},$$

$$\begin{pmatrix} e(2q+1) \\ e(2q+2) \\ e(2q+3) \\ e(2q+L) \\ \hline e(2q+L+1) \\ e(2q+L+2) \\ e(3q+L) \end{pmatrix} \tilde{\mathbf{W}}_3 \mathbf{x}_3 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ \hline 1 & 0 & 0 \\ -w(1) & 1 & 0 \\ -w(2) & -w(1) & 1 \end{pmatrix} \begin{pmatrix} x(2q+L+1) \\ x(2q+L+2) \\ x(3q+L) \end{pmatrix}$$

und das Endergebnis resultiert aus:

$$\hat{\mathbf{x}}_{u,b} = -(\tilde{\mathbf{W}}_u^H \tilde{\mathbf{W}}_u)^{-1} \left[\tilde{\mathbf{W}}_u^H (\tilde{\mathbf{W}}_1 \mathbf{x}_1 + \tilde{\mathbf{W}}_3 \mathbf{x}_3) \right]. \quad (2.44)$$

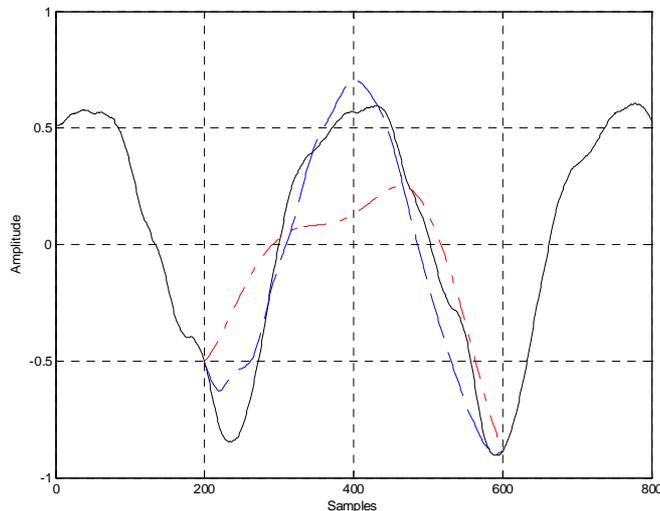


Abb. 2-13: Vergleich der beiden Interpolationsmethoden,
 durchgezogen/schwarz: Zielsignal,
 strichpunktiert/rot: Gesamtsystem,
 strichliert/blau: reduziertes System.

Eine Analyse von Gesamtsystem und reduzierter Variante zeigt den Vorteil letzterer in Bezug auf die interpolierte Signalfolge. Die durchgezogene/schwarze Linie in Abb. 2-13 zeigt einen Ausschnitt von 800 Samples (ca. 20ms) aus einem Musiksignal. Die Interpolation von 400 Punkten (Samples 201 bis 600) bei Betrachtung des Gesamtsystems (Fall (a), Abb. 2-10) ist strichpunktiert/rot dargestellt, während die reduzierte Variante (Fall (b), Abb. 2-10) die strichlierte/blau Kurve ergibt.

Ein objektiver Vergleich der beiden Systeme ist durch die Berechnung des Signal-Rausch-Abstandes (SNR) innerhalb des interpolierten Ausschnitts möglich. Dabei wird der Fehler der Interpolation (Zielsignal – Schätzung) als Rauschkomponente verwendet. Zur Messung eines mittleren SNR dienen verschiedene Musiksignale, in denen mehrere Ausfallszeitpunkte betrachtet werden. Das in Abb. 2-14 dargestellte Ergebnis zeigt eine Verbesserung des mittleren SNR von ca. 1.14dB.

Ein wesentlicher Aspekt der Systemminimierung liegt zudem in der Verringerung der Rechenkomplexität.

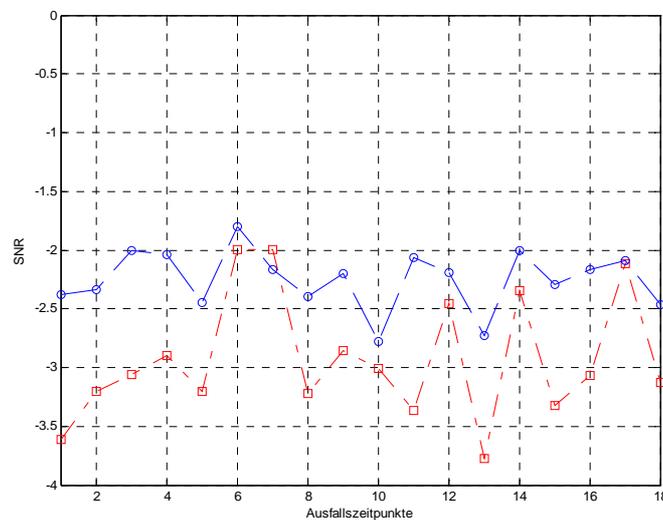


Abb. 2-14: SNR der interpolierten Signale zu verschiedenen Ausfallszeitpunkten
 strichpunktirt/rot: Gesamtsystem
 strichliert/blau: reduziertes System

2.4.4. Adaption der Interpolation in E-DROCON

Der Einsatz der Interpolation in E-DROCON basiert auf der Idee, mithilfe der dem Signal inhärenten Periodizität ein zeitliches Muster in der Signalvergangenheit zu detektieren und in den zukünftigen Ausfall einzufügen, um damit eine Interpolation zu ermöglichen. Der entsprechende Algorithmus umfasst die Schritte:

- Detektion einer Periodizität im Signal,

- Kopieren des passenden Ausschnitts aus dem Puffer und Einfügen an entsprechender Stelle im Ausfall,
- Definition des Interpolationsblocks,
- Schätzung der Filterkoeffizienten,
- Ermittlung der fehlenden Signalsamples.

Die Detektion der optimalen Signalfolge im Puffer wird mithilfe der Substitutionsmethode realisiert. Anschließend wird dieser Teil an die Stelle des aktuellen, ausgefallenen Blocks sowie einer bestimmten Anzahl an nachfolgenden Blöcken kopiert. Abb. 2-15 zeigt das Zustandekommen des Interpolationsblocks. Die Länge ist als variabler Parameter definiert und kann von der Ausfallscharakteristik – vorausgesetzt diese ist bekannt – abgeleitet werden. Ein passender Wert kann mithilfe der Häufigkeitsverteilung der vorkommenden Ausfalllängen bestimmt werden. Ist bei einer gegebenen Übertragungstrecke die Wahrscheinlichkeit am größten, dass ein Ausfall z. B. sechs Millisekunden dauert, so kann die Länge des Interpolationsblocks vorab ebenfalls auf 6ms eingestellt werden. Die Anpassung (Verkürzung oder Verlängerung) an die tatsächliche Länge des Ausfalls erfolgt durch die Überprüfung des Statusbits. Eine Verkürzung ist notwendig, wenn das fehlerfrei übertragene Signal wieder verfügbar ist, bevor der gesamte Interpolationsblock verwendet wurde. Bei anhaltendem Ausfall ist eine Verlängerung erforderlich, d. h. es werden weitere Signalpunkte von der entsprechenden Stelle aus der Vergangenheit in den aktuellen fehlenden Block kopiert.

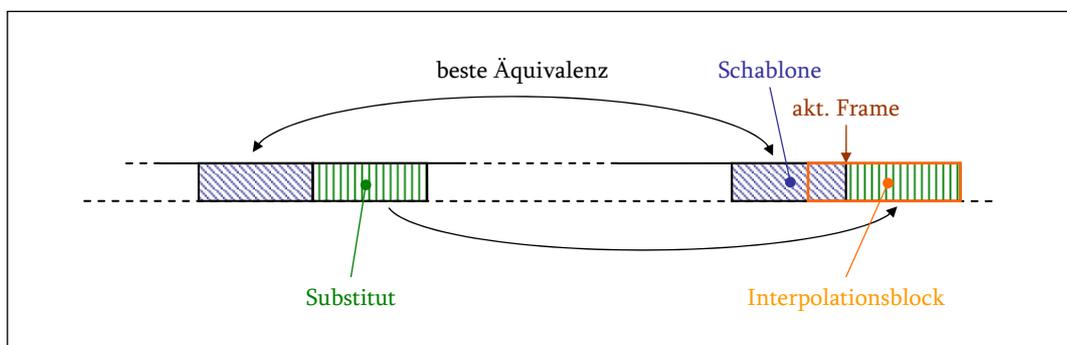


Abb. 2-15 Interpolation in E-DROCON

2.4.5. Pufferverbesserung

Die im Puffer gespeicherten Daten dienen zur Schätzung des Zielsignals während eines Ausfalls. Je nachdem, in welchem zeitlichen Abstand diese Ausfälle auftreten, kann der Puffer entweder vollständig mit originalen oder zu einem Teil mit geschätzten Signalpunkten gefüllt sein, wenn in diesem Teil ein früherer Ausfall stattgefunden hat. Die aktuelle Schätzung ist damit nicht mehr ausschließlich vom tatsächlichen Signal, sondern auch von einigen bereits geschätzten Werten abhängig, was in Folge die Qualität der Verschleierung vermindert. Die Pufferverbesserung versucht nun durch eine rückwirkende Modifikation den Anteil der nachgebildeten Daten zu minimieren. Die zusätzlichen Verarbeitungsschritte führen dazu, die dem Ausfall nachfolgenden originalen Datensamples in voller Größe im Puffer auszunutzen und werden aus diesem Grund am Ende des Überblendungsbereichs zwischen geschätztem und originalem Zielsignal ausgeführt.

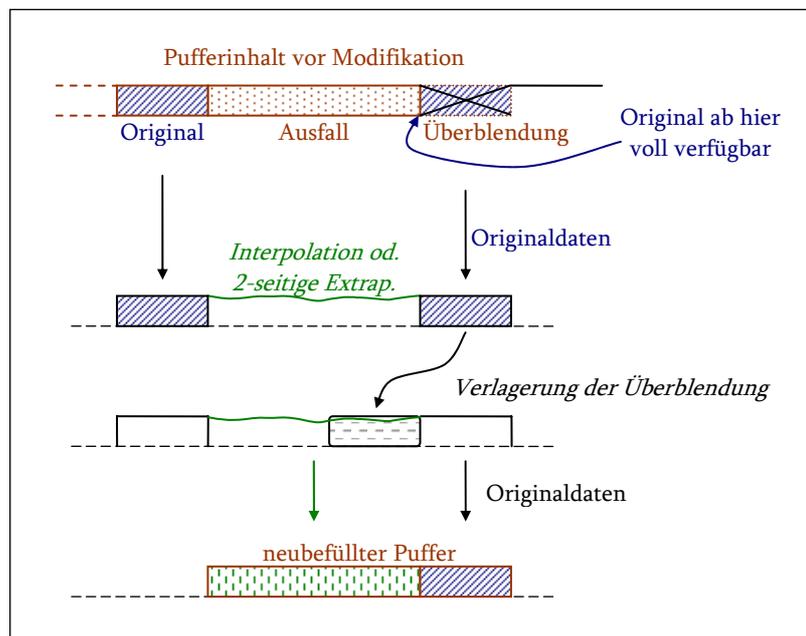


Abb. 2-16: Prinzip der Pufferverbesserung

In Abb. 2-16 ist das Prinzip schematisch dargestellt: Vor und nach dem Ausfall wird jeweils ein Ausschnitt des Zielsignals (Original) bestimmt. Mit diesen beiden Vektoren wird nun entweder eine zweiseitige Extrapolation mit anschließender Überblendung (nach [Etter 1996]) oder eine Interpolation (nach [Vaseghi 1996]) durchgeführt. Dadurch wird der

Überblendungsbereich, der vor der Modifikation direkt im Anschluss an den Ausfall, d. h. an der Stelle der ersten wieder verfügbaren originalen Zielsamples lag, in den Ausfall hinein verschoben. Der ursprüngliche Puffer wird teilweise neu befüllt, wobei vor allem die Ausnutzung der (ungefensterten) Samples des wiedervorhandenen Zielsignals eine Verbesserung des Pufferinhalts bewirken.

2.5. Zusammenfassung

Die Implementierung der Ausfallsverschleierung in einem Monosignal stützt sich auf die klassischen Zeitbereichsverfahren Extrapolation und Substitution. Erstere eignet sich vor allem zur Überbrückung sehr kurzer Ausfälle ($\leq 3\text{ms}$), bei längeren Blöcken des Datenverlusts führt der Energieabfall im geschätzten Zielsignal zu hörbaren Amplitudenschwankungen und damit zur Verminderung der Audioqualität. Der wesentliche Vorteil der Extrapolation liegt darin, dass ein phasenkontinuierlicher Übergang zwischen Zielsignal und dessen Nachbildung gewährleistet werden kann, ohne dafür eine Pufferung im aktiven Signalpfad zu benötigen. Dies kann mit der Substitutionsmethode aufgrund der Nichtstationarität des Eingangssignals nicht erreicht werden, da es nur bei exakt periodischen Signalen möglich ist, mittels Periodizitätsschätzung einen Signalblock in der Vergangenheit zu detektieren, der vollkommen ident mit dem ausgefallenen ist. Bei längeren Signalunterbrechungen erweist sich die Substitution hingegen als effizientere Verschleierungstechnik.

In der Interpolation lassen sich die Stärken beider Methoden kombinieren, allerdings ist für deren Realisierung die Verfügbarkeit von Signalpunkten vor und nach dem Ausfall zwingend notwendig. Um diese Forderung im E-DROCON-Modul zu erfüllen, werden an die Stelle der unbekannt zukünftigen Samples passende, aus dem Signalpuffer entnommene, eingesetzt. Die spezielle Beschränkung der Dimension der Datenvektoren und –matrizen bei der Ermittlung der Filterkoeffizienten einerseits sowie bei der Erzeugung der

ausgefallenen Signalblöcke andererseits führt zu einem rechenzeitoptimierten Gesamtalgorithmus.

Zusammenfassend werden hier nochmals die Rechenschritte der vier Hauptzweige aufgelistet:

(1) Erster Block eines Signalausfalls:

- Periodizitätsschätzung nach Gl. (2.25),
- Bestimmung des Ausschnitts aus dem Puffer und "Einfügen" desselben nach dem aktuellen Block (d. h. in die Zukunft),
- Schätzung der Filterkoeffizienten nach Gl. (2.43),
- Beginn der Interpolation, d. h. Schätzung des Zielsignals im aktuell ausgefallenen Block nach Gl. (2.44).

(2) Mittlerer Block:

- Fortführung der Schätzung mit den gespeicherten Filterkoeffizienten nach Gl. (2.44).

(3) Erster Block des Zielsignals nach einem Ausfall:

- Weiterführung der Schätzung (wie vorhin),
- Initialisierung des Bereichs zur Überblendung zwischen Schätzungs- und Zielsignal,
- Speicherung des Zielsignals (ohne überblendeter Schätzung) in separatem Puffer zur Pufferverbesserung.

(4) Ausfallsfreier Zustand:

- Kontrolle, ob Überblendung vollständig durchgeführt; wenn ja: "Durchschleifen" des Zielsignals durch die Verschleierungseinheit; wenn nein: Weiterführung des geschätzten Signals und der Überblendung,
- weiteres Befüllen des separaten Puffers, bis dieser gefüllt ist; Durchführung der Pufferverbesserung.

Kapitel 3

Perzeptive Evaluierung der einkanaligen Verschleierung

Zur Evaluierung der einzelnen Algorithmen wurden mehrere subjektive Testreihen durchgeführt, welche auf Richtlinien der "International Telecommunication Union", Sektion "Radiocommunication" (ITU-R) basieren. Recommendation ITU-R BS. 1116-1 betrachtet Methoden zur subjektiven Bewertung von geringfügigen Signalverschlechterungen in Audiosystemen und beinhaltet neben einer standardisierten Testmethode Empfehlungen zur Auswertung und Dokumentation der Datensätze [ITU-R 1997].

Die gezielte Untersuchung von geringfügigen Signalverschlechterungen ist speziell auf zwei Anwendungsgebiete ausgerichtet. Einerseits werden Verfahren zur Datenreduktion, beispielsweise Audiokodierverfahren, evaluiert, siehe [Thiede 2000] und darin angeführte Literatur, andererseits dient der Standard als Basis für die Untersuchung von Methoden zur Fehler- und Ausfallsverschleierung von Signalübertragungssystemen. Subjektive Tests, die in ähnlicher Weise wie die nachfolgend erläuterten Hörtests durchgeführt werden, sind unter anderem in [Wasem 1988], [Liao 2001] oder [Esquef 2006] beschrieben.

Im Allgemeinen sind Hörversuche mit einem beträchtlichen Arbeits- und Zeitaufwand verbunden. Deshalb wurden objektive Testverfahren als Soft- oder Hardwarelösungen entwickelt, deren Funktionsweise ebenfalls in Empfehlungen der ITU-R dokumentiert sind

([ITU-R 1998]) bzw. auch in [Thiede 2000] näher erläutert werden. Die nach dem gehörangepassten Messverfahren PEAQ (*perceptual evaluation of audio quality*) arbeitende objektive Methode ermöglicht eine ausgedehnte Evaluierung mit einer großen Anzahl an Testsignalen, ohne dadurch Arbeits- bzw. Zeitressourcen drastisch zu erhöhen.

Für die geplanten Tests wurde die auf dem PEAQ-Algorithmus basierte Software OPERA (<http://www.opticom.de>) verwendet. Im Verlauf der objektiven Untersuchungen stellte sich jedoch heraus, dass die Software in der vorliegenden Anwendung nicht zielführend eingesetzt werden kann. Der Grund liegt darin, dass das Verfahren eine Mittelung über die gesamte Länge des Audiobeispiels vornimmt, die weder ausgeschaltet noch vom Anwender auf einen gewünschten Bereich eingestellt werden kann. Da die im vorliegenden Anwendungsfall angenommenen Ausfälle einen geringen Anteil (ca. 2%) des gesamten Testsignals betragen, wird das objektive Bewertungsergebnis durch die Mittelung verfälscht, d. h. die perzeptive Auswirkung der Verschleierungsverfahren wird nicht korrekt beurteilt.

In der Entwurfsphase des experimentellen Tests muss darauf geachtet werden, dass dieser ausschließlich die Effekte der unabhängigen Faktoren sichtbar macht und Mehrdeutigkeiten durch unkontrollierte Faktoren verhindert werden. Weiters muss beim Design berücksichtigt werden, dass die Versuchspersonen (VPn) nicht überfordert werden, d. h. die an sie gestellte Aufgabe sollte klar verständlich formuliert und einfach zu lösen sein.

3.1. Testvorbereitung

3.1.1. Auswahl der Versuchspersonen

Die ITU-R empfiehlt bei der Bewertung von geringfügigen Verschlechterungen von Audiosignalen ausschließlich auf Personen mit Expertise in der Detektion von geringfügigen Signalverschlechterungen ("*expert listeners*") zurückzugreifen. Da für die vorliegenden

Testreihen kein entsprechendes Panel zur Verfügung stand, wurde die Untersuchung mit nichtgeschulten Versuchspersonen durchgeführt. Informelle Tests während der Vorbereitung zeigten allerdings, dass diese Personengruppe mithilfe einer adäquaten Trainingsphase imstande war, eine kritische Bewertung der Signalverschlechterungen vorzunehmen. Zudem dienen die Hörversuche in erster Linie dem Vergleich der vorgeschlagenen Algorithmen, d. h. sie sollen einen groben Überblick über deren Leistungsverhalten geben. Eine Gegenüberstellung mit anderen Techniken der Ausfallsverschleierung ist im Rahmen der vorliegenden Arbeit nicht vorgesehen.

Die Anzahl der VPn jeder Testreihe wurde ebenfalls nicht primär von den in [ITU-R 1997] genannten Kriterien abhängig gemacht, sondern richtete sich nach der Verfügbarkeit der Personen. Verfahren zur Vor- bzw. Nachselektion (*pre-screening*, *post-screening*) der VPn wurden zum Teil berücksichtigt.

3.1.2. Testmethode

Als Testmethode wurde das in psychoakustischen Tests üblicherweise verwendete Verfahren *double-blind triple-stimulus with hidden reference* eingesetzt. Dabei werden einer Versuchsperson hintereinander drei akustische Stimuli A, B und C vorgespielt, wobei A als Referenzstimulus definiert wird und im vorliegenden Testdesign dem fehlerfreien Zielsignal (d. h. einem Signal gänzlich ohne Ausfälle bzw. verschleierte Signalteilen) entspricht. Einem der Stimuli B oder C wird zufällig das mit (künstlich eingebauten) Ausfällen behaftete und mittels Verschleierungstechnik wiederhergestellte Signal zugewiesen ("Objektstimulus" oder "Objekt"), der verbleibende Stimulus wird nochmals mit dem Referenzsignal belegt und als versteckte Referenz bezeichnet. Mithilfe einer in [ITU-R 1997] bzw. [ITU-R 2003] standardisierten fünfteiligen Bewertungsskala wurde die VP angewiesen, jeweils den Unterschied zwischen B und A sowie C und A zu beurteilen. Den fünf Ankerpunkten der Skala sind die Bezeichnungen "nicht wahrnehmbar", "wahrnehmbar, aber nicht störend",

"leicht störend", "störend" und "sehr störend" zugewiesen, siehe Tab. 3-1. Die Skala selbst ist kontinuierlich¹⁴ ausgeführt, d. h. Werte zwischen den Ankerpunkten sind zulässig.

Verschlechterung	subjektiver Grad
nicht wahrnehmbar	5.0
wahrnehmbar, aber nicht störend	4.0
leicht störend	3.0
störend	2.0
sehr störend	1.0

Tab. 3-1: ITU-R fünfteilige Bewertungsskala, [ITU-R 1997] & [ITU-R 2003]

Üblicherweise erfolgt die Analyse der Daten auf Basis des subjektiven Differenzgrades (SDGs), der als Differenz der Bewertungen von Objektstimulus und versteckter Referenz definiert ist:

$$SDG = G_{Objekt} - G_{verst.Referenz}$$

Dabei wird davon ausgegangen, dass die VP in der Lage ist, die versteckte Referenz korrekt zuzuordnen, folglich nimmt der SDG einen Wert zwischen 0 und -4 ein. Da die vorliegenden Testreihen nicht mit *expert listeners* durchgeführt wurden, wäre die Verwendung des SDGs nicht optimal gewesen – falsche Detektionen der versteckten Referenz durften nicht ausgeschlossen werden. Abschnitt 3.4 beschreibt die Auswertung der Testdaten und beinhaltet eine genaue Diskussion der möglichen Bewertungsfehler.

Vergleichbare subjektive Testreihen, wie sie z. B. in [Esquef 2006] beschrieben sind, bezeichnen die Bewertung der VPn auch als *mean opinion score* (MOS) und definieren hierfür einen unteren Grenzwert. Liegt der MOS im Mittel über einem Wert von 3.5, so wird die Qualität der evaluierten Signalverarbeitung als gut empfunden.

¹⁴ Im vorliegenden Testdesign erfolgte die Bewertung mit einem Schieberegler.

3.1.3. Testmaterial, Audiosignale

Die Selektion der Audiobeispiele unterliegt laut ITU-R keinen speziellen Kriterien, die Empfehlungen beziehen sich ausschließlich auf die Evaluierung von Signalverschlechterungen bei Komprimierungs- und Kodierungsalgorithmen und sind für die qualitative Untersuchung der vorgeschlagenen Verschleierungsmethoden ungeeignet. In der ersten Testreihe wurde versucht, eine breite Palette durch verschiedene Musikgenres, Soloinstrumente und Sprache herzustellen. Für die Auswahl in den nachfolgenden Testreihen wurden die Ergebnisse der ersten berücksichtigt (z. B. weitere Untersuchung von Signalen, die im ersten Versuch mäßige bis schlechte Qualität der Verschleierung ergaben).

Da Mittel- und Langzeitgedächtnis als unzuverlässige Entscheidungsträger bei psychoakustischen Tests angesehen werden, sollte sich die Testprozedur ausschließlich auf das Kurzzeitgedächtnis stützen. Dementsprechend sollte die Länge der Exzerpte so kurz wie möglich festgelegt werden. Unter Berücksichtigung der Simulation einer bestimmten Anzahl von Signalausfällen im Bereich bis zu zehn Millisekunden je Beispiel, wurde eine Dauer von fünf Sekunden als akzeptabel befunden. Die Gesamtlänge aller Ausfälle je Exzerpt wird in Prozent der Dauer definiert, die exakte Ausfallsstatistik variiert in jeder Testreihe und wird im nachfolgenden Abschnitt genau spezifiziert.

Eine Session sollte nicht länger als 20 bis 30 Minuten dauern, durch Ermüdung lässt die Konzentration der VPn nach. Im aktuellen Testdesign mussten die einzelnen Trials jeweils vom Probanden selbst gestartet werden, die automatische Abfolge der Stimuli mit A-B / A-C war fix vorgegeben und eine Wiederholung eines Stimulus (oder mehrerer Stimuli) war nicht möglich. Die Pausen zwischen den Stimuli sollten wiederum sehr kurz sein, jedoch dürfen etwaige Maskierungen oder andere Artefakte (unsauberes Ein- bzw. Ausblenden) den Höreindruck in keiner Weise verändern. Obwohl das Tempo der Testprozedur von der VP abhängig ist, kann eine grobe Relation zwischen Dauer einer Session und Anzahl der darin enthaltenen Trials hergestellt werden.

Allgemein waren sämtliche Audiobeispiele als Monosignale gespeichert, deren Wiedergabe erfolgte über Kopfhörer. Der Test wurde von den VPn einzeln durchgeführt. Eine individuelle Abhaltung der Tests bietet den Vorteil, dass sich die Probanden in Ruhe mit dem Testmaterial beschäftigen können und nicht durch andere Personen etc. gestört werden.

3.2. Durchführung der Testreihen

Die den Testreihen zugrunde liegenden unabhängigen Faktoren sind jeweils darauf ausgerichtet, spezifische Merkmale der Algorithmen zu extrahieren. Da eine genaue Aufklärung der Probanden über Details (Art, Anzahl, etc.) der unabhängigen Faktoren nicht notwendig war, durften innerhalb eines Versuchsdurchlaufs mehrere getestet werden, ohne die Probanden in deren Entscheidungsverhalten zu verwirren oder zusätzlich zu belasten. Beispielsweise wurden in der ersten Testphase als unabhängige Faktoren unterschiedliche

(1) Algorithmen,

(2) Audiobeispiele sowie

(3) Ausfallsszenarien

gewählt, in Trials bestehend aus den drei Stimuli A, B, C eingebaut und diese vollständig randomisiert aneinandergereiht. Die VPn wurden instruiert, jeden Trial in gleicher Weise zu beurteilen. Der Bewertungsprozess war zweigeteilt: einerseits war die Zuweisung von Objekt und versteckter Referenz zu den Stimuli B und C notwendig, andererseits musste der Grad der Verschlechterung des Objektes in Bezug auf den Referenzstimulus eingeschätzt werden.

Zur Überprüfung der Reliabilität der VPn waren in der ersten Testreihe zwei gleichartige Sessions eingebaut, die sich lediglich in der Reihenfolge der Trials unterschieden. Der Vergleich der Antworten beider Sessions wurde in einer Art *post-screening*-Prozedur verwertet. Da das Panel der Versuchspersonen bereits beim ersten Hörversuch festgelegt wurde, war eine Reliabilitätsprüfung in den nachfolgenden Tests nicht mehr notwendig.

Grundsätzlich bestand jede Testreihe aus einer Trainingphase und dem eigentlichen Hörtest. Die Trainingsphase diente dazu, die Versuchspersonen mit der Testumgebung vertraut zu machen. Zum einen wurden die verwendeten Audiobeispiele in ihrer Originalform vorgestellt, zum anderen spielte das Kennenlernen der Artefakte, welche die Verschleierungsalgorithmen erzeugen, eine wesentliche Rolle. Der Vergleich zu Signalen, die mit Totalausfällen behaftet sind, war ebenfalls für jedes Beispiel möglich. Speziell im Einführungsteil der ersten Testreihe wurde die Beschreibung der Testmethode, der Bewertungsskala und des Bewertungsprozesses ausführlich mit jeder Versuchsperson diskutiert.

In die anschließende Bewertungsphase (tatsächlicher Hörtest) wurden Kontrollbedingungen inkludiert. Typischerweise sind dies Trials, die aus dem Referenzstimulus A und zwei versteckten Referenzstimuli in B und C bestehen (Typ I). Zusätzlich wurden Kontrolltrials verwendet, deren Objekt einem mit Totalausfällen behafteten Signal entspricht (Typ II). Dadurch sollte während des Hörtests die Orientierung an den Grenzen der Bewertungsspannbreite 5.0 bzw. 1.0 stabilisiert werden.

3.3. Spezifikationen der einzelnen Testreihen

3.3.1. Testreihe 1: Extrapolation vs. Interpolation

Die erste Testreihe diente zur Evaluierung der beiden Verschleierungsmethoden "Extrapolation" und "Interpolation". Die jeweiligen Parameterwerte für Pufferlänge, Filterordnung, etc. wurden gleich eingestellt.

Die sechs Audiobeispiele repräsentieren einen der unabhängigen Faktoren und setzen sich zusammen aus:

- Auszüge aus den drei Stilrichtungen Pop, Jazz und Klassik,

- je einem Sample der Instrumente Klavier und Geige (solistisch),
- einem Sprachbeispiel: weibliche Stimme (deutsch).

Als weiterer unabhängiger Faktor wird der simulierte Übertragungsfehler in Form einer Ausfallstatistik definiert: 2% der Gesamtlänge eines Beispiels sind durch einen Signalausfall gestört bzw. ungültig und müssen auf Empfängerseite mithilfe der Verschleierungseinheit rekonstruiert werden.

Die Verteilung der 17 Ausfälle wird folgendermaßen festgelegt:

Länge eines Ausfalls [ms]	Anzahl
1	2
2	2
3	1
4	2
5	2
6	1
7	2
8	2
9	1
10	2

Tab. 3-2: Verteilung der Ausfälle von Testreihe 1

Da sich die spezielle Signalcharakteristik zum Ausfallszeitpunkt (Spektrum, Stationarität) ebenfalls auf den Erfolg der Verschleierungstechnik auswirkt, werden für jedes Audiobeispiel vier unterschiedliche Ausfallsszenarien zufällig generiert. Diese ergeben den dritten unabhängigen Faktor.

Die Berücksichtigung aller Faktoren resultiert in einer Gesamtmenge von $2 \cdot 6 \cdot 4 = 48$ Trials, hinzu kommen noch vier Kontrolltrials (jeweils zwei des Typs I bzw. II). Würde die gleiche Wiedergabereihenfolge der Trials bei allen Probanden beibehalten, so kann nicht mit Sicherheit festgestellt werden, ob die subjektiven Urteile tatsächlich aufgrund der Qualität der Stimuli getroffen wurden, oder ob vielmehr die spezielle Sequenz einen Einfluss auf das Antwortverhalten hat. Aus diesem Grund ist es notwendig, die Reihenfolge der Trials für jede VP zufällig zu generieren. Die Randomisierung kann auf zwei Arten erfolgen, entweder

über den gesamten Beispielsatz oder innerhalb der Subgruppe jedes Faktors [ITU-R 1997]. Erstere Methode wird im Allgemeinen angewandt, wenn eine homogene Verteilung der Faktoren vorausgesetzt werden kann, d. h. wenn diese unabhängig voneinander sind und keine unterschiedlichen Einflüsse auf das Antwortverhalten der VPn haben. In Situationen, in denen die Faktoren eine Rangordnung wie beispielsweise unterschiedliche Schwierigkeitsgrade oder eine ähnliche Form der Abhängigkeit besitzen, kann keine Homogenität erwartet werden. Deshalb dürfen die Trials der einzelnen Subgruppen nicht vermischt werden, sondern die Randomisierung darf nur innerhalb jeder Gruppe durchgeführt werden.

In den drei vorliegenden Testreihen werden ausschließlich unabhängige Faktoren untersucht und folglich eine vollständige Randomisierung aller Trials für jede VP realisiert. Zusätzlich bietet diese Variante den Vorteil, dass Eingewöhnungsphasen zu Beginn (die trotz Einführungsphase nicht vollständig ausgeschlossen werden können) und Ermüdungserscheinungen gegen Ende des Hörtests keine systematischen Verzerrungen der Testresultate bewirken. Die Zuweisung von Objekt und versteckter Referenz zu den Stimuli B und C erfolgt ebenfalls zufällig, sie ist auch dem Leiter des Experiments unbekannt. Somit ist eine (un-)bewusste Beeinflussung ausgeschlossen.

Die Trainingsphase dieser Testreihe setzt sich aus der Vorstellung der Testumgebung beziehungsweise dem Kennenlernen der Audiobeispiele zusammen. Im ersten Teil erhält die Versuchsperson eine Einführung über Inhalt und Ziel des Tests, dessen genauen Ablauf (inklusive Erläuterung der Bewertungsskala) sowie die Bedienung des Testprogramms. Anschließend können die originalen, mit Ausfällen behafteten und ausfallsverschleierte Audiobeispiele in beliebiger Reihenfolge (wiederholt) angehört werden.

Wie im vorangegangenen Abschnitt beschrieben, besteht der eigentliche Hörtest aus zwei Sessions, um die Retest-Reliabilität zu prüfen.

3.3.2. Testreihe 2: Parametersetting Extrapolation

In der zweiten Testreihe wird die Optimierung des Parametersettings der Extrapolation untersucht. Die Auswertung der Testreihe 1 (siehe Vergleich V3, Abschnitt 3.4) zeigt, dass die Extrapolationstechnik in Bezug auf den Faktor "Audiobeispiel" die schlechteste Performanz bei den Samples von Geige und Sprache liefert. Folglich wird geprüft, ob durch eine Änderung der Parameter des Algorithmus eine Verbesserung der Qualität der verschleierten Signale erzielt werden kann. Zusätzlich werden zwei weitere Instrumentenbeispiele – Saxofon und Gitarre – verwendet.

Der zweite unabhängige Faktor wird durch drei unterschiedliche Parametersettings gebildet, in denen die Länge des Puffers (zwischen 600 und 1000 Samples) sowie die Ordnung des Prädiktionsfehlerfilters (zwischen 100 und 300 Samples) variiert wird.

Die Ausfallsstatistik (2% der Länge des Exzerpts ist ausfallsbehaftet) wird beibehalten, allerdings werden die Ausfälle wie folgt angenommen:

Länge eines Ausfalls [ms]	Anzahl
2	9
3	6
4	5
5	4
6	3

Tab. 3-3: Verteilung der Ausfälle von Testreihe 2 und Testreihe 3

Aufgrund der größeren Gesamtanzahl (27) ergibt sich eine gleichmäßigere Verteilung der Ausfälle innerhalb des Audiobeispiels, somit ist die Beschränkung auf ein Ausfallszenario zulässig.

Insgesamt ergeben sich für den zweiten Hörversuch $4 \cdot 3 = 12$ Trials, welche wiederum um vier Kontrolltrials erweitert werden. Letztere sind einerseits vom Typ I, andererseits werden zwei neue Trials generiert, die den Vergleich mit einem simplen Verschleierungsverfahren ermöglichen. Dabei wird jeder Signalausfall durch gefärbtes Rauschen (Rosa Rauschen mit

TP-Filterung bei 8kHz) mit an die Signalleistung vor dem Ausfall angepasster Amplitude substituiert (Typ III).

Die Trainingsphase beschränkt sich auf das Kennenlernen der neuen Samples. Sechs Subjekte mussten eingeschult werden, da sie für die erste Testreihe nicht zur Verfügung standen.

Als Folgerung der positiven Reliabilitätsprüfung in Testreihe 1 wird die Durchführung einer einzigen Session je VP als ausreichend angesehen.

3.3.3. Testreihe 3: Periodizitätsschätzung

Die dritte Testreihe zielt darauf ab, verschiedene Methoden der Periodizitätsschätzung zu evaluieren. Eine Technik wurde bereits in Testreihe 1 bei der Interpolation verwendet, die beiden Alternativen werden speziell für diese Untersuchung entwickelt.¹⁵

Die Auswahl der Audiobeispiele beschränkt sich aufgrund der Erkenntnisse der vorigen Tests auf die Samples von Jazz, Klassik und Geige. Die Ausfallstatistik wird von Testreihe 2 übernommen, jedoch werden für diesen Versuch zwei Ausfallszenarien generiert. Dadurch ergeben sich insgesamt $3 \cdot 3 \cdot 2 = 18$ Trials plus je zwei Kontrolltrials vom Typ I und II.

Eine Trainingsphase ist nicht mehr notwendig, da alle Versuchspersonen bereits an der ersten Testreihe teilgenommen haben.

3.4. Statistische Analyse

Die statistische Analyse der Datensätze der drei subjektiven Testreihen erfolgt durch den Vergleich von Stichproben hinsichtlich ihrer zentralen Tendenz, welche abhängig vom Messniveau und der Häufigkeitsverteilung der Daten gewählt werden muss. Als Messniveau

¹⁵ Eine detaillierte Beschreibung der drei Techniken findet sich in Kapitel 2.

gilt die fünfteilige Bewertungsskala (vgl. Tab. 3-1), die aufgrund der Charakterisierung der Stufen durch Adjektive, wie z. B. "wahrnehmbar" oder "störend" etc., der Klasse der Ordinalskalen zuzuordnen ist.

Vor der eigentlichen Stichprobenanalyse muss die Gültigkeit der Daten überprüft und Fehlbewertungen der Subjekte separiert bzw. als Sonderfall behandelt werden. Dazu kann eine Einteilung der subjektiven Bewertungen dienen, wie sie in Tab. 3-4 vorgenommen wird. In Fall 1 entspricht Stimulus C tatsächlich der versteckten Referenz, Stimulus B beinhaltet den Objektstimulus. Das Subjekt erkennt allerdings einen Unterschied zwischen (Referenz-) Stimulus A und Stimulus C und identifiziert letzteren somit als Objektstimulus. Äquivalent dazu beschreibt Fall 2 die umgekehrte Variante, dass C tatsächlich das Objekt enthält, obwohl die Versuchsperson die Zuordnung B = Objekt vornimmt. Im dritten Fall werden beide Stimuli B und C als Objekt bewertet, unabhängig von der tatsächlichen Zuordnung, auch wenn dies theoretisch nicht möglich ist.¹⁶ Fall 4 ist als Sonderfall zu behandeln, da (unabhängig von der tatsächlichen Zuordnung) das Subjekt nicht imstande ist, den Objektstimulus zu identifizieren, d. h. einen Unterschied zwischen B und A oder zwischen C und A wahrzunehmen. Diese Situation führt zu folgender Schlussfolgerung: Die perzeptive Auswirkung der Verschleierung der Ausfälle ist unerschwellig und kann daher nicht detektiert werden. Dieser Fall darf streng genommen nicht als Fehlbewertung bezeichnet werden.

		Zuordnung der Stimuli	
		$B_{St} = 5.0 / C_{St} < 5.0$	$B_{St} < 5.0 / C_{St} = 5.0$
Bewertung durch VP	$B_{VP} = 5.0 / C_{VP} < 5.0$	korrekt	Fall 1
	$B_{VP} < 5.0 / C_{VP} = 5.0$	Fall 2	korrekt
	$B_{VP} < 5.0 / C_{VP} < 5.0$	Fall 3	
	$B_{VP} = 5.0 / C_{VP} = 5.0$	Fall 4	

Tab. 3-4: Möglichkeiten der Fehlbewertung

¹⁶ Dieser Fall tritt trotz Einführungsphase und Erklärung des Bewertungsvorganges vereinzelt auf.

Bei der im Folgenden für jede Testreihe getrennt durchgeführten Analyse werden die Daten gruppiert:

- G1 ... gültige Daten: Die subjektive Zuordnung entspricht der tatsächlichen,
- G2 ... falsche Detektion: Zusammenfassung der Fälle 1 bis 3 aus Tab. 3-4,
- G3 ... B0/C0 Detektion: Verschleierung nicht wahrnehmbar (Fall 4 aus Tab. 3-4).

3.4.1. Testreihe 1: Extrapolation vs. Interpolation

Der erste Hörversuch, bestehend aus zwei Sessions, wurde mit 40 Probanden durchgeführt.

In Tab. 3-5 sind die unabhängigen Faktoren zusammengefasst.

Faktor, F_i	Abkürzung	Umfang, N_i
F1: Audiobeispiel	[aud]	$N_1: 6$
F2: Algorithmus	[alg]	$N_2: 2$
F3: Ausfallsszenario	[dos]	$N_3: 4$
F4: Session	[run]	$N_4: 2$
F5: Versuchsperson	[sub]	$N_5: 40$

Tab. 3-5: Unabhängige Faktoren von Testreihe 1

Abb. 3-1 zeigt die Verteilung des korrigierten Datensatzes, welcher die Gruppen G1 und G3 umfasst; die falschen Detektionen (G2) werden nicht berücksichtigt. Die kontinuierliche Bewertungsskala wird im Histogramm in 0.5er Schritte eingeteilt. Die Verteilungsfunktion der Daten entscheidet unter anderem darüber, welches statistische Verfahren zum Vergleich der Stichproben herangezogen werden muss. Konkret wird zwischen normalverteilten und nicht normalverteilten Daten unterschieden. Das Histogramm ermöglicht eine einfache grafische Überprüfung und zeigt für den vorliegenden Datensatz eine asymmetrische Verteilungsfunktion.¹⁷

¹⁷ Zusätzlich können rechnerische Methoden zur Prüfung auf Normalverteilung vorgenommen werden. Eine Möglichkeit bietet der Chi-Quadrat-Test, der als Prüfgröße die Ähnlichkeit zwischen gegebener Verteilung und Normalverteilung bestimmt, [Bortz 2005], [Zöfel 1992]. Mittelwert und Standardabweichung der Normalverteilung müssen dabei nicht spezifiziert sein, da der Test auf der Beurteilung von Asymmetrie (Schiefe) und Kurtosis der gegebenen Verteilung beruht. Als Ergebnis liefert der Test die Aussage, ob bei einem Signifikanzlevel von 5% die Nullhypothese – die gegebene Verteilung entspricht einer Normalverteilung – verworfen werden muss oder beibehalten werden kann.

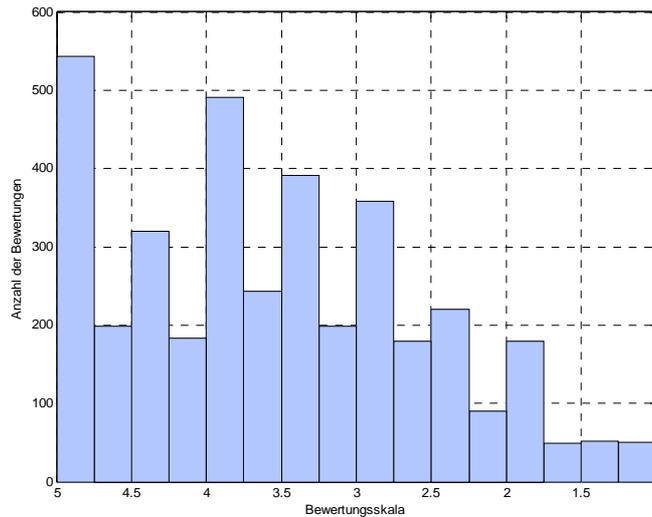


Abb. 3-1: Verteilung der subjektiven Bewertung, Testreihe 1
(Abszisse: 5 ... wahrnehmbar, aber nicht störend,
1 ... sehr störend)

Die deutlichen Anhäufungen an den Ankerpunkten und etwas abgeschwächt in der Mitte der Stützstellen deuten darauf hin, dass sich die VPn relativ stark an der Skaleneinteilung orientiert haben. Diese Tatsache lässt sich mit der in Abb. 3-2 dargestellten grafischen Oberfläche des Testprogramms erklären.

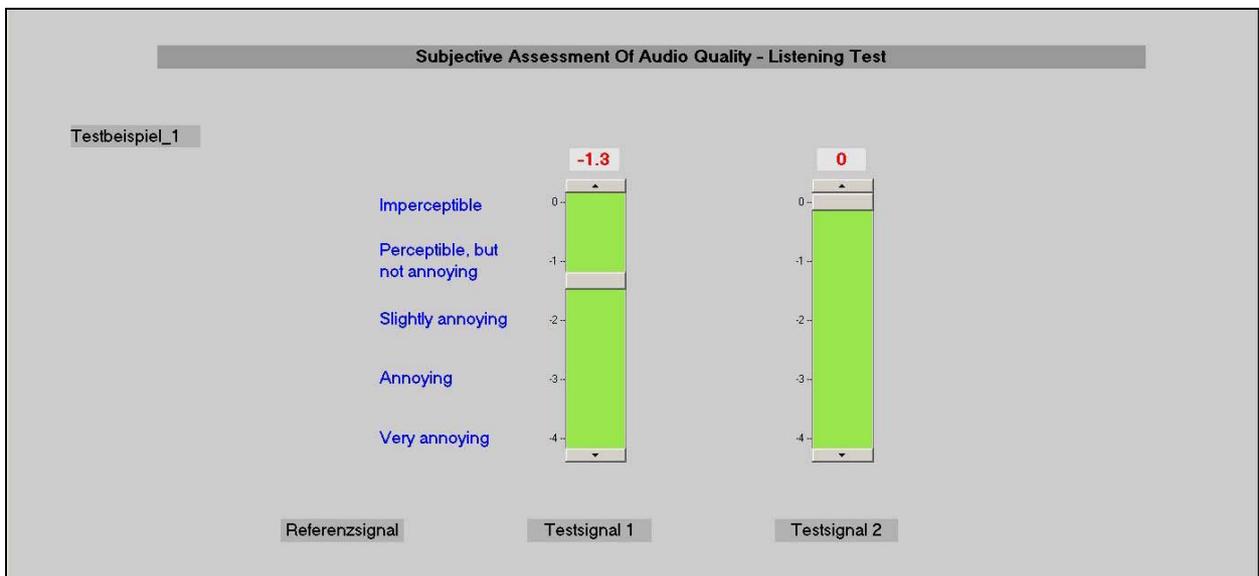


Abb. 3-2: Grafische Oberfläche des subjektiven Testprogramms

Die rechnerische Überprüfung der ersten Testreihe bestätigt, dass keine Normalverteilung der Daten vorliegt. Aus diesem Grund darf der arithmetische Mittelwert nicht als Messwert

für den Vergleich von Stichproben herangezogen werden. Ein gültiges Maß für nicht normalverteilte Daten stellt der Medianwert dar. Das Verfahren zur Überprüfung von Unterschiedshypothesen muss in Abhängigkeit des jeweiligen Messwertes gewählt werden. Entsprechend empfiehlt sich im vorliegenden Fall die Verwendung parameterfreier Prüfverfahren, wobei der U-Test von Mann und Whitney, [Mann 1947], [Bortz 2005], zum Vergleich von zwei unabhängigen Stichproben dient. Liegen mehr als zwei unabhängige Stichproben vor, so wird der H-Test nach Kruskal und Wallis, [Kruskal 1952], [Zöfel 1992], zum Vergleich herangezogen. Beide Testmethoden sind für den vorliegenden Datensatz notwendig – die unabhängigen Faktoren Algorithmus und Session bestehen aus jeweils zwei Stichproben, bei den restlichen ist der Stichprobenumfang größer als zwei.

Vergleich V1: Prüfung auf signifikanten Unterschied zwischen den Sessions

Mit dieser Überprüfung wird die Reliabilität der VPn untersucht. Die Nullhypothese für den U-Test lautet: Es besteht kein signifikanter Unterschied (auf dem 5%-Signifikanzlevel) zwischen den beiden Sessions. Die Auswertung des U-Tests erfolgt durch die Transformation der Prüfgröße U in den auf einer Standardnormalverteilung basierenden z-Wert. Letztendlich zeigt dieser z-Wert, dass die Nullhypothese nicht verworfen werden darf, sondern ihre Gültigkeit behält. Konkret bedeutet das, dass die Reproduzierbarkeit der Antworten der Probanden signifikant ist.

Da die Versuchspersonen (bis auf wenige Ausnahmen) bereits durch Testreihe 1 offensichtlich ein klares Verständnis für die Aufgabenstellung entwickelt hatten, war eine Reliabilitätsprüfung in den weiteren Testreihen nicht mehr notwendig. Infolgedessen werden die restlichen zwei Testreihen auf je eine Session begrenzt.

Vergleich V2: Prüfung auf signifikanten Unterschied zwischen den Algorithmen

Dieser Vergleich stellt das Hauptziel der ersten Testreihe dar. Wie in Vergleich V1 wird die Prüfung mittels U-Test vorgenommen, wobei sich die beiden Stichproben nun durch Faktor F₂ unterscheiden. Die Nullhypothese – es besteht kein signifikanter Unterschied zwischen den Algorithmen – darf nicht verworfen werden, sie ist sogar sehr signifikant (1%-Signifikanzlevel).

Somit ist die Qualität beider Verschleierungstechniken als gleichwertig anzusehen. Infolgedessen wurde mithilfe einer zweiten Testreihe versucht, die dem Algorithmus inhärenten Parameter zu optimieren, um ein verbessertes Leistungsverhalten zu erzielen.

Vergleich V3: Prüfung auf signifikanten Unterschied zwischen den Audiobeispielen

Der Faktor [aud] führt zu einer Unterteilung des Datensatzes in 6 Stichproben, welche mit dem H-Test ausgewertet werden. Im Gegensatz zum U-Test basiert der H-Test nicht auf dem standardnormalverteilten z-Wert, sondern wird entsprechend einer Chiquadrat-Verteilung transformiert. Der resultierende Chiquadrat-Wert bestimmt die Wahrscheinlichkeit, mit der die aufgestellte Nullhypothese (= kein Unterschied zwischen den Audiobeispielen) zugunsten der Gegenhypothese verworfen werden darf.

Erstmals verliert die Nullhypothese ihre Gültigkeit, da die Medianwerte der einzelnen Audiobeispiele signifikant voneinander abweichen. Der Chiquadrat-Wert liefert keine Details über die Unterschiede, diese können mithilfe des Boxplots dargestellt werden, siehe Abb. 3-3.

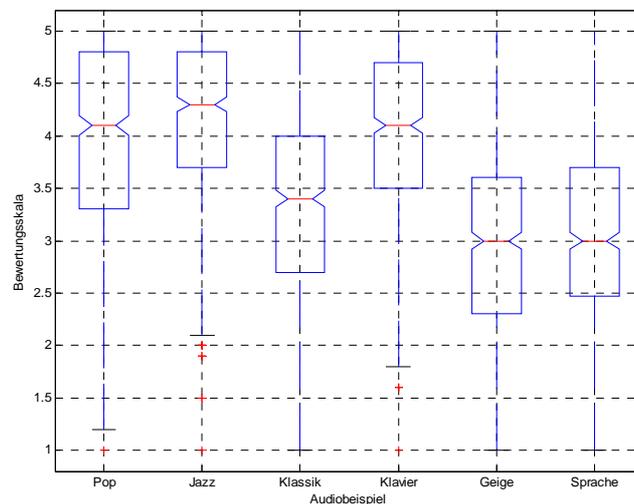


Abb. 3-3: Vergleich der Audiobeispiele, Testreihe 1

Die Grafik bestätigt die ursprüngliche Annahme, dass die Art des Musik- (bzw. Sprach-) signals wesentlichen Einfluss auf das Leistungsverhalten des Algorithmus hat. Dabei können grob zwei Klassen unterschieden werden: Bei den Beispielen Pop, Jazz und Klavier kann die

Qualität der Verschleierung als gut (der Medianwert liegt oberhalb der Stufe "hörbar, aber nicht störend") eingestuft werden, wohingegen die Samples Klassik, Geige und Sprache etwas problematischer zu verarbeiten sind (mit Medianwerten bei "leicht störend").

Vergleich V4: Prüfung auf signifikanten Unterschied zwischen den Ausfallsszenarien

Die letzte noch verbleibende Überprüfung, die sich auf nur einen Parameter bezieht, ist der Vergleich der vier Ausfallsszenarien. Entsprechend werden die vier Stichproben mit dem H-Test ausgewertet, welcher auf einem Signifikanzniveau von 5% die Gegenhypothese als richtige Behauptung bestätigt.

Diese Aussage bedeutet, dass der Ausfallszeitpunkt innerhalb eines Audiobeispiels wesentlichen Einfluss auf die Qualität der Verschleierungsmethode hat, wobei beide Algorithmen wiederum gleiche Leistung bei gleichem Szenario aufweisen.

Fehlbewertungen und Kontrolltrials

In Abb. 3-4 ist der prozentuelle Anteil der Fehlbewertungen je Audiobeispiel dargestellt. Die Aufteilung in Gruppe G2 und G3 zeigt, dass die Anzahl der tatsächlichen Falschdetektionen im Verhältnis zu den B0/C0 Detektionen viel geringer ist. Zudem ist auch in dieser Grafik die 2er-Klassifizierung der Audiobeispiele erkennbar. Sie führt zur selben Schlussfolgerung wie in Vergleich V3: Der Anteil der Fehlbewertungen ist bei den Samples von Pop, Jazz und Klavier deutlich höher als bei denen von Klassik, Geige und Sprache. Folglich ist bei ersteren die Verschleierung von Ausfällen weniger kritisch als bei letzteren.

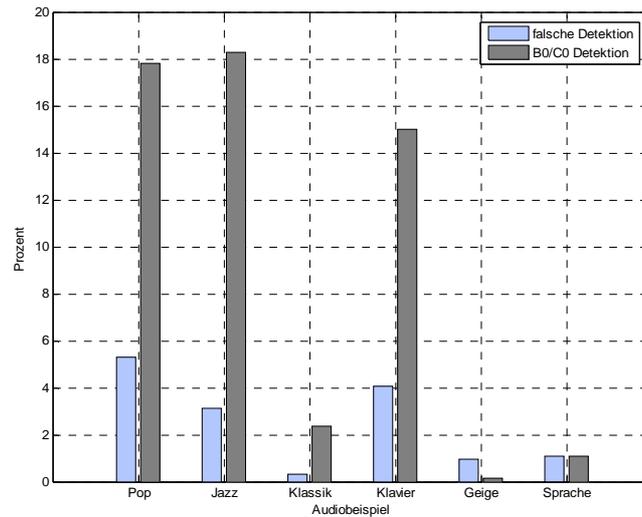


Abb. 3-4: Fehlbewertung von Testreihe 1

Die vier Kontrolltrials wurden willkürlich den Audiobeispielen Pop, Klavier (Typ I: Referenzstimulus in B und C) sowie Jazz und Klassik (Typ II: Ausfall, keine Verschleierung) zugeordnet, siehe Abb. 3-5. Es ist ersichtlich, dass sich Typ I sehr gut an der oberen Grenze der Bewertungsskala platziert. Bei Typ II ist die Positionierung an der unteren Grenze nicht so extrem ausgeprägt, im Vergleich zur Bewertung der "normalen" Trials der entsprechenden Audiobeispiele liegen sie jedoch eindeutig darunter.

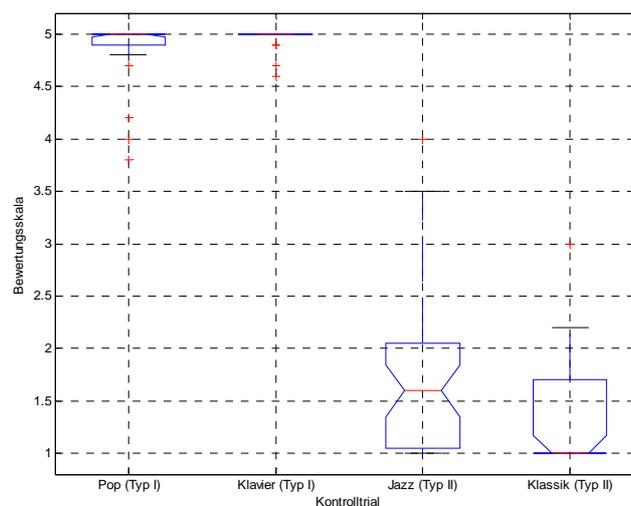


Abb. 3-5: Kontrolltrials von Testreihe 1

Sämtliche Auswertungen von Testreihe 1 sind in Anhang A, Abschnitt A.2 zusammengefasst.

3.4.2. Testreihe 2: Parametersetting Extrapolation

Mit der zweiten Testreihe soll geprüft werden, inwieweit eine Optimierung eines der beiden Algorithmen möglich ist. Im Hinblick auf die Implementierung in einer Echtzeitumgebung wurde der Extrapolationsalgorithmus evaluiert, da dieser geringere Rechenkomplexität besitzt. Einen Überblick über die unabhängigen Faktoren, deren Details in Abschnitt 3.3 erläutert sind, bietet Tab. 3-6. Die Verteilung der Daten sowie die vollständige Auswertung der zweiten Testreihe findet sich in Anhang A, Abschnitt A.3; Vergleich V5 liefert die wesentliche Aussage der Testreihe.

Faktor, F_i	Abkürzung	Umfang, N_i
F ₁ : Audiobeispiel	[aud]	N ₁ : 4
F ₂ : Parametersetting	[par]	N ₂ : 3
F ₃ : Versuchsperson	[sub]	N ₃ : 15

Tab. 3-6: Unabhängige Faktoren von Testreihe 2

Vergleich V5: Prüfung auf signifikanten Unterschied zwischen den Parametersettings

Die Prüfung mittels H-Test zeigt, dass die Unterschiede zwischen den Parametersettings signifikant sind (siehe Abb. 3-6) und folglich eine gezielte Adjustierung der Pufferlänge und der Filterordnung das Leistungsverhalten des Algorithmus entsprechend verbessert. Folgende Werte wurden eingestellt: (1) 600/120 Samples, (2) 800/200 Samples und (3) 1000/280 Samples. Der Vergleich von (1) und (2) zeigt, dass zu geringe Werte zu schlechteren Ergebnissen führen. Zwischen (2) und (3) ist jedoch keine signifikante Abweichung mehr festzustellen. Deshalb ist eine Begrenzung nach oben hin in Bezug auf die Reduzierung des Rechen- und Speicheraufwands sinnvoll.

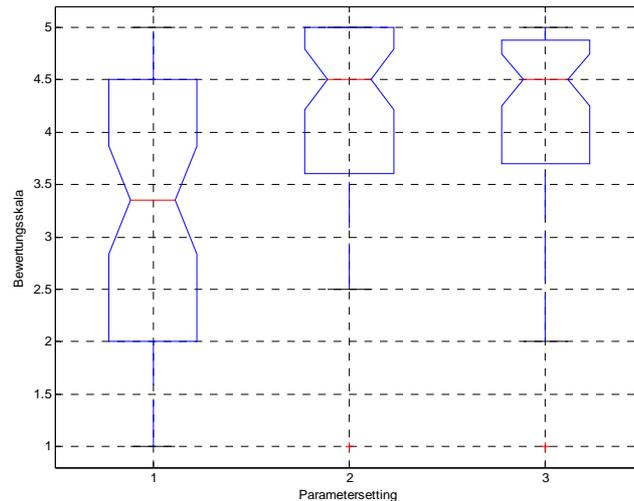


Abb. 3-6: Vergleich der Parametersettings, Testreihe 2

3.4.3. Testreihe 3: Periodizitätsschätzung

Die dritte Testreihe dient zur Evaluierung eines speziellen Teils des Interpolationsalgorithmus, der für die Schätzung der Periodizität im Signal verantwortlich ist. Dabei wurden die in Kapitel 2 vorgeschlagenen drei Techniken (Nullstellenanalyse, Differenzfunktion nach Goodman bzw. Differenzfunktion nach de Cheivgne) miteinander verglichen. In Tab. 3-7 sind die unabhängigen Faktoren zusammengefasst, deren Beschreibung beinhaltet Abschnitt 3.3. Details zur Auswertung sind in Anhang A, Abschnitt A.4 zu finden, nachfolgender Vergleich veranschaulicht den Zusammenhang der Stichproben von Faktor F_2 und liefert eine Aussage über die Qualität der Techniken.

Faktor, F_i	Abkürzung	Umfang, N_i
F_1 : Audiobeispiel	[aud]	N_1 : 4
F_2 : Periodizitätsschätzung	[per]	N_2 : 3
F_3 : Ausfallsszenario	[dos]	N_3 : 2
F_4 : Versuchsperson	[sub]	N_4 : 21

Tab. 3-7: Unabhängige Faktoren von Testreihe 3

Vergleich V6: Prüfung auf signifikanten Unterschied zwischen den Algorithmen zur Periodizitätsschätzung

Das Ergebnis des H-Tests deutet auf einen signifikanten Unterschied zwischen den Algorithmen hin. Zur genaueren Verifizierung sind in Abb. 3-7 die jeweiligen Boxplots der drei Methoden dargestellt. Hieraus lässt sich erkennen, dass der Unterschied zwischen der Nullstellenanalyse (1) und der Differenzfunktion nach Goodman (2) signifikant ist, während de Cheveigne's Version der DF (3) weder zum ersten noch zum zweiten Algorithmus eine deutliche Abweichung aufweist.

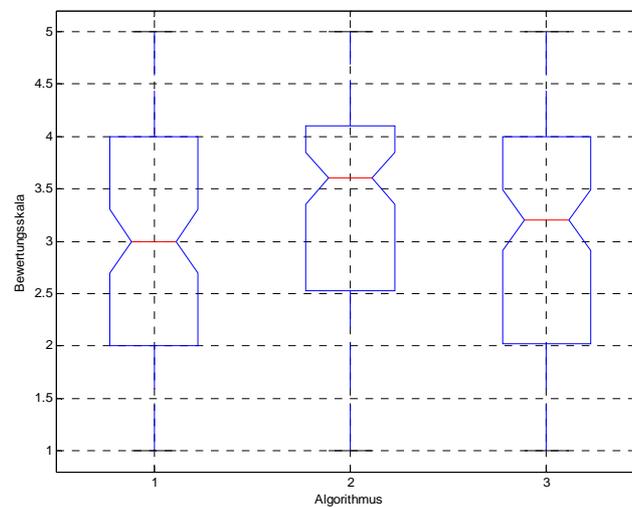


Abb. 3-7: Vergleich der Algorithmen zur Periodizitätsschätzung, Testreihe 3
 (1) Nullstellenanalyse
 (2) DF nach Goodman
 (3) DF nach de Cheivgne

3.5. Zusammenfassung

Das Kapitel beschreibt die perzeptive Evaluierung der Audioqualität der fehlerverschleierte Signale, die mithilfe von drei subjektiven Testreihen durchgeführt wurde. Die Testreihen

unterscheiden sich geringfügig im Design und zielen jeweils darauf ab, bestimmte unabhängige Faktoren zu untersuchen:

- Testreihe 1: Algorithmus, Audiobeispiel, Ausfallsszenario;
- Testreihe 2: Parametersetting, Audiobeispiel;
- Testreihe 3: Art der Periodizitätsschätzung, Audiobeispiel, Ausfallsszenario.

Das Ergebnis der statistischen Auswertung aller Testdaten ist in Tab. 3-8 zusammengefasst, wobei die Signifikanzprüfung auf dem 5%-Niveau basiert. Die Bezeichnung "teilweise signifikant" weist darauf hin, dass die einzelnen Samples eines Faktors sowohl signifikante als auch nicht signifikante Unterschiede zueinander aufweisen. Als Ergebnis liefert der Signifikanztest die Aussage, dass mindestens zwischen zwei Samples ein signifikanter Unterschied detektierbar ist. Die genauen inneren Zusammenhänge können mithilfe der jeweiligen Boxplots, welche in Anhang A dargestellt sind, verifiziert werden.

Testreihe	unabh. Faktor	Unterschied
1	Algorithmus Audiobeispiel Ausfallsszenario	nicht signifikant teilw. signifikant teilw. signifikant
2	Parametersetting Audiobeispiel	teilw. signifikant teilw. signifikant
3	Periodizitätssch. Audiobeispiel Ausfallsszenario	teilw. signifikant signifikant signifikant

Tab. 3-8: Zusammenfassung der statistischen Auswertung der drei Testreihen

Kapitel 4

Mehrkanalige Systeme

In der digitalen Audiotechnik repräsentieren mehrkanalige Anordnungen den Stand der Technik bei Aufnahme- und Wiedergabesystemen. Dementsprechend werden Übertragungstrecken von den Quellen zu den Senken ebenfalls mit mehreren zeitlich synchronisierten, parallelen Kanälen ausgeführt. Bei Konzertveranstaltungen und Bühneninstallationen reichen Mehrkanalanordnungen von Stereomikrofonie über unterschiedliche Varianten der Hauptmikrofonierung für Surroundaufnahmen (wie beispielsweise OCT Surround, Decca Tree, Hamasaki Square, etc. [Neumann 2001]) bis zu verschiedenen Formen der Stützmikrofonie. Konzertante Szenarien repräsentieren das Hauptanwendungsgebiet für mehrkanalige Ausfallsverschleierungsverfahren.

Diese Form der Audioübertragung findet aber auch in Studios über verschiedene physikalische Layer statt (z. B. Lichtwellenleiter, AES-EBU, CAT5), wobei aus verschiedenen Gründen, wie etwa Synchronisationsverlust, Ausfälle auftreten können. In kritischen Anwendungen wie z. B. im Sendebetrieb einer Rundfunkanstalt sind derartige Fehler unbedingt durch Sicherheitseinrichtungen mit geringer Verarbeitungslatenz zu kompensieren.

Audioübertragung im Internet ist zwar weniger zeitkritisch als die oben genannten Bereiche, dafür treten Übertragungsfehler häufiger auf und die damit verbundene Reduktion der Signalqualität ist deutlich stärker. Die Verbesserung des *quality-of-service* mithilfe einer

Ausfallsverschleierung, wie sie in dieser Arbeit vorgeschlagen wird, ist in jedem Fall von Vorteil. Eine interessante Anwendungsmöglichkeit im Rahmen der Netzwerkübertragung beschreiben Sinha et. al. als *distributed immersive musical performance*, einer Realisierung eines gemeinsamen Konzerts von Musikern, die örtlich voneinander getrennt sind, über das Internet [Sinha 2003]. Die zeitliche Synchronisation der Instrumentalisten sowie die Bewahrung der perzeptiven Charakteristika des (virtuellen) Konzertsaaes für alle Künstler und Zuhörer kennzeichnen die wesentlichen Anforderungen an ein solches System. Demnach kommt insbesondere die ultrakurze Latenzzeit der vorgeschlagenen Ausfallsverschleierung zu tragen.

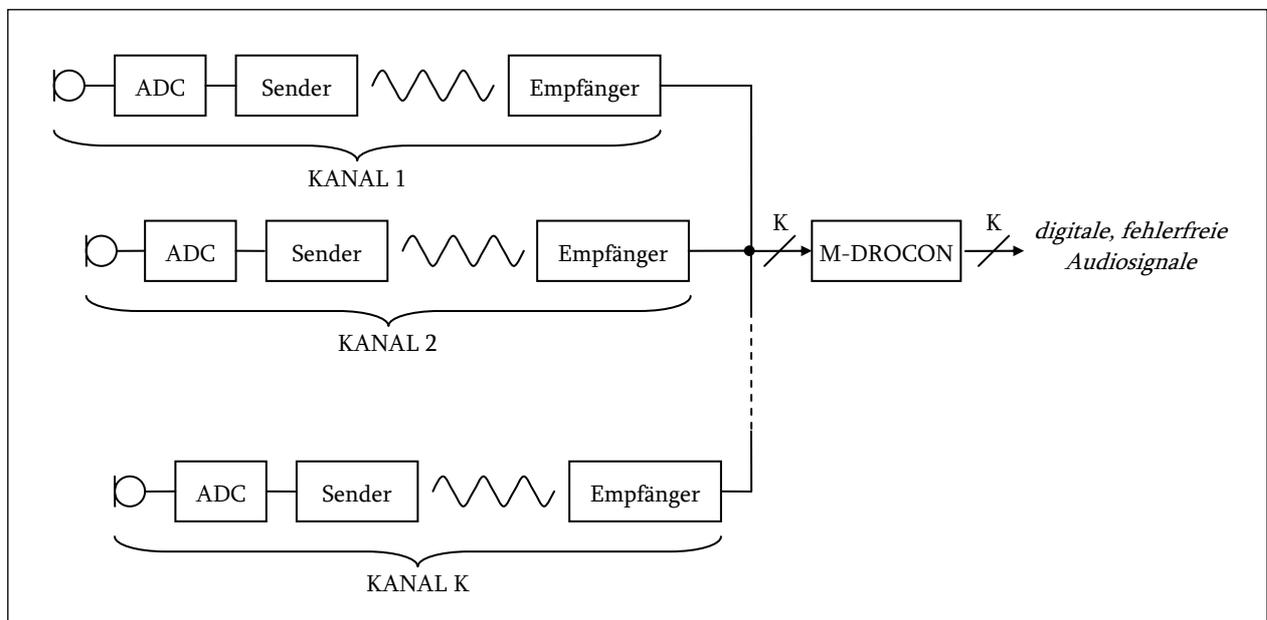


Abb. 4-1: Mehrkanalige Übertragungsstrecke

Selbstverständlich ist es in einer mehrkanaligen Anordnung möglich, jeden Kanal getrennt zu behandeln und folglich K einzelne E-DROCON-Einheiten zu implementieren. Im vorliegenden Kapitel werden jedoch Verfahren aufgezeigt, die in der Lage sind, die parallelen Kanäle zur Ausfallsverschleierung gemeinsam zu nutzen. Kennzeichnend für mehrkanalige Techniken ist die Strategie, einen ausgefallenen Kanal aus einem Kanal oder mehreren alternativen Kanälen zu rekonstruieren. Das System aus Abb. 1-1 wird um $(K - 1)$ Übertragungspfade erweitert, wie in Abb. 4-1 schematisch dargestellt. Die Funktionalität der

entwickelten Techniken hängt primär von folgenden Kriterien der Signale in den Kanälen ab:

- (1) Kreuzkorrelation zwischen den Kanälen: sie beschreibt den statistischen Zusammenhang zwischen bzw. die gemeinsame Signalinformation in den Kanälen. Dabei spielen die räumlichen Distanzen zwischen den Signalaufnehmern (Mikrofonen) sowie die Entfernungen zu den jeweiligen Quellen (Soloinstrumente, Instrumentengruppen) eine entscheidende Rolle.
- (2) Kohärenzfunktion: lässt sich aus der Kreuzkorrelation ableiten und misst den Anteil der linear abhängigen Energieanteile zweier Signale. Sie stellt ein probates Maß für die Größe des statistischen Zusammenhangs der Kanäle einer Mehrkanalanordnung dar.

4.1. Stand der Forschung

Einige einkanalige Verfahren können für die Anwendung in einem Mehrkanalsystem adaptiert werden. Der Übergang von der Intra-Kanal-Verschleierung (*intra-channel concealment*) auf die Inter-Kanal-Verschleierung (*inter-channel concealment*) [Sinha 2003] wird nachfolgend für die Verschachtelungsmethode (*interleaving*), diverse Substitutionstechniken sowie die lineare Prädiktion erläutert.

Bei mehrkanaligen Systemen muss berücksichtigt werden, dass bei einem Totalausfall der gesamten Übertragungsstrecke keine mehrkanalige Verschleierung mehr möglich ist, d. h. es sollte in jedem Fall die Option einer Intra-Kanal-Variante vorgesehen werden.

Sinha et. al. [Sinha 2003] erweitern die Substitutionsmethoden aus [Goodman 1986] bzw. [Liao 2001] um mehrere parallele Kanäle und implementieren die Mustererkennung (*pattern matching*) bzw. die Replikation des Signalverlaufs mithilfe der Tonhöhenenerkennung (*pitch waveform replication*) in Form von Inter-Kanal-Verschleierungsstrategien. Die zwischen den Übertragungskanälen stattfindende Substitution führt zu einer Interpolation der fehlenden

Signalanteile. Dazu ist allerdings ein spezielles Übertragungsformat erforderlich, weshalb das Verfahren zur Klasse der sender- und empfängerseitig basierten Systeme gehört. Dieses fußt auf dem Prinzip der Verschachtelung und sieht eine Zusammenfassung unterschiedlicher Kanäle in den Paketen eines Übertragungskanals vor. Folglich wird nicht für jedes zu übertragende Audiosignal ein eigener Kanal zur Verfügung gestellt. Die Zusammensetzung der Pakete erfolgt unter Berücksichtigung der Korrelation zwischen den Signalen. Bei rein empfängerseitig basierten Techniken, wie sie in der vorliegenden Arbeit untersucht werden, ist jedenfalls ein Kanal je Audiosignal notwendig.

In aufwändigeren Verfahren ist vor der eigentlichen Ausfallsverschleierung eine Stufe zur Signalaufbereitung vorgesehen, welche die zeitdiskreten Audiosignale in eine für die nachfolgende Verarbeitung optimierte Repräsentation bringt. Wiese und Sedlmeyer beschreiben zum Beispiel in [Wiese 2002] ein Verfahren, das eine Subbandaufteilung des gesamten breitbandigen Audiosignals voraussetzt. Bei Ausfall werden die einzelnen Subbänder getrennt untersucht und eine Verschleierung von einzelnen Spektralkomponenten bzw. Subbändern durchgeführt. Die Teilsignale können dabei sowohl im Zeit- als auch im Frequenzbereich vorliegen.

Bruhn ([Bruhn 2005a], [Bruhn 2005b], [Bruhn 2006]) beschreibt in mehreren als Patentschriften vorliegenden Arbeiten ein mehrkanaliges Verschleierungsverfahren, das auf dem Prinzip der Filterung eines fehlerfreien Signals zur Erzeugung einer Schätzung des ausgefallenen Signals beruht. Im Detail behandelt werden stereofone Anordnungen sowie mehrkanalige Szenarien bestehend aus einem Hauptkanal und mehreren Seitenkanälen, wobei im Hauptkanal alle Signalanteile der Seitenkanäle enthalten sein müssen. Dementsprechend ist für die Verschleierung eines Seitenkanals grundsätzlich lediglich der Hauptkanal erforderlich, umgekehrt müssen jedoch bei Ausfall des Hauptkanals alle Seitenkanäle fehlerfrei übertragen werden. Die zweigeteilte Technik setzt sich aus einer Analyse- und einer Synthesestufe zusammen und verwertet im Gegensatz zu den Substitutionstechniken nicht direkt das Zeitsignal, sondern ein davon abgeleitetes parametrisches Modell. Äquivalent zur Extrapolation im Einkanalfall erfolgt die Schätzung

der Parameter während der fehlerfreien Übertragung aller Kanäle. Bei Ausfall in einem Kanal wird mit der entsprechenden Syntheseinheit aus den Parametern die Schätzung des Zielsignals vorgenommen, wofür konkret die Verwendung von adaptiven Filtern vorgeschlagen wird. In der einfachsten Realisierung werden die letzten gültigen Filterkoeffizienten konstant gehalten und mit dem jeweils aktuellen Block des Substitutionssignals kombiniert, um das Zielsignal im jeweils aktuellen Block zu schätzen. Da die Koeffizienten eine Repräsentation der Korrelation von Ziel- und Substitutionssignal sind, wird aufgrund der nichtstationären Signale das Filter mit Fortdauer des Ausfalls immer ungenauer. Als Verbesserung wird vorgeschlagen, das geschätzte Zielsignal wieder an den Eingang des adaptiven Filters zurückzuführen und damit die Schätzung der Koeffizienten vorzunehmen. Schließlich wird noch die Kombination dieser beiden Versionen angeführt, wobei die aktuelle Schätzung des Zielsignals aus einer Gewichtung der letzten gültigen Koeffizienten sowie jenen, die aus den vergangenen geschätzten Datenpunkten gewonnen werden, erzeugt wird.

In der vorliegenden Arbeit wurden adaptive Filter mit dem *least mean square-* (LMS-) bzw. *normalised least mean square-* (NLMS-) Algorithmus auf ihre Eignung als Verschleierungstechniken untersucht. Beide Methoden basieren auf der Ausnutzung der Kreuzkorrelation zwischen Ziel- und Substitutionskanal. In einer allgemeinen Mehrkanalanordnung von Audiosignalen ist die Varianz der Kreuzkorrelation zwischen zwei Kanälen im Normalfall sehr hoch und kann zudem abhängig von den geometrischen Gegebenheiten relativ gering sein. Um eine ausreichende Adaption des Filters zu erreichen, sind lange Adaptionzeitkonstanten notwendig. Andererseits widerspricht die Nichtstationarität der Signale dieser Forderung, d. h. es wird im Allgemeinen eine möglichst schnelle Adaptierung des Filters aufgrund des Signalverlaufs gewünscht. Eine Verbesserung des Konvergenzverhaltens bietet die Methode der affinen Projektion (AP), die die Fehlerminimierung in einem mehrdimensionalen Raum beschreibt. Die AP kann als Erweiterung des NLMS-Algorithmus betrachtet werden, da die Schätzung der Filterkoeffizienten nicht mehr von einem skalaren Wert, sondern von einem Fehlervektor abhängig ist. Umgekehrt kann der NLMS-Algorithmus als Sonderfall der affinen Projektion

erster Ordnung interpretiert werden. Die Steigerung der Konvergenzrate hat eine Erhöhung des Rechenaufwandes zur Folge, weshalb speziell schnelle Implementierungsvarianten im Zeit- und Frequenzbereich entwickelt wurden, siehe u. A. [Haykin 2002], [Gay 1993], [Gay 1995], [Rupp 1998] und [Rombouts 2000].

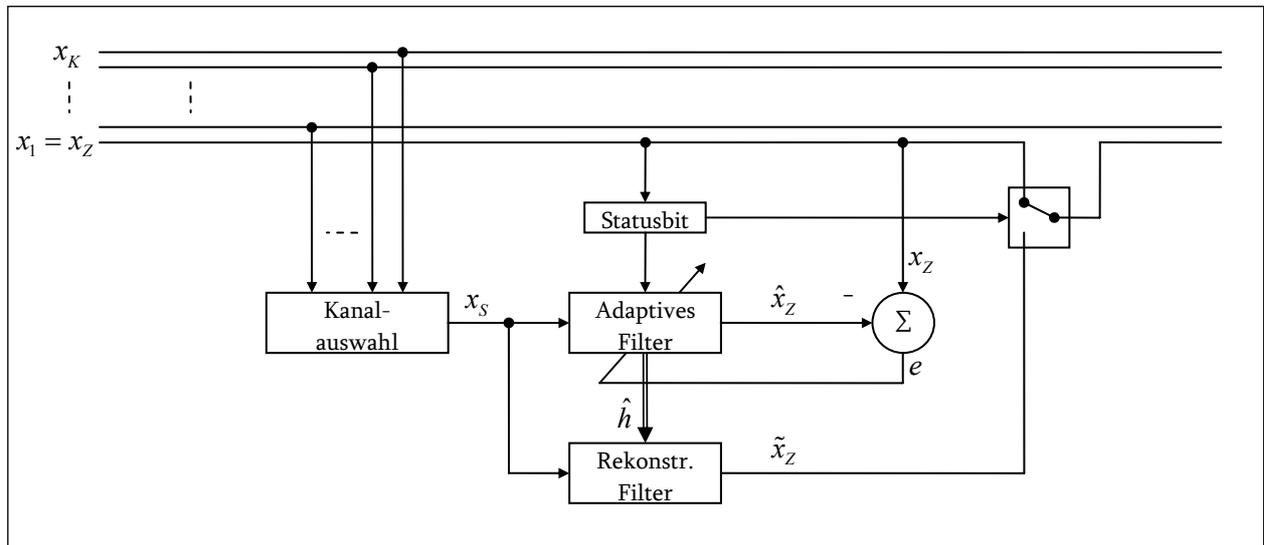


Abb. 4-2: Ausfallsverschleierung mittels adaptivem Filter

Das Prinzip der Ausfallsverschleierung mittels adaptiver Filter ist in Abb. 4-2 für einen mit Ausfällen behafteten Kanal dargestellt. Eine Auswahleinheit sucht aus allen zur Verfügung stehenden Kanälen unter Berücksichtigung ihrer Fehlerfreiheit und Korrelation zu dem ausfallsbehafteten Signal (wird in der Abbildung als Zielsignal \mathbf{x}_z bezeichnet) die besten Kandidaten und berechnet daraus ein Substitutionssignal \mathbf{x}_s , das am Eingang des adaptiven Filters anliegt. Die geschätzten Filterkoeffizienten $\hat{\mathbf{h}}$ werden laufend aktualisiert, solange das Zielsignal nicht durch einen Ausfall gestört ist. Mit dem NLMS-Algorithmus lassen sich die aktuellen Koeffizienten wie folgt berechnen:

$$\hat{\mathbf{h}}(n+1) = \hat{\mathbf{h}}(n) + \mu \frac{\mathbf{x}_s(n)}{\|\mathbf{x}_s(n)\|^2 + \nu} e(n),$$

mit der Adaptionkonstante μ und einem Stabilisierungsparameter ν . Der Sampleindex wird mit n bezeichnet. Das Fehlersignal wird sampleweise berechnet mit

$$e(n) = x_z(n) - \hat{x}_z(n),$$

wobei \hat{x}_z die laufende Schätzung des Zielsignals darstellt,

$$\hat{x}_z(n) = \mathbf{x}_s(n)^T \hat{\mathbf{h}}(n).$$

In Analogie dazu lauten die Beziehungen für den AP-Algorithmus:

$$\begin{aligned} \hat{\mathbf{x}}_z(n) &= \mathbf{X}_s^T(n) \hat{\mathbf{h}}(n) \\ \mathbf{e}(n) &= \mathbf{x}_z(n) - \hat{\mathbf{x}}_z(n) \\ \hat{\mathbf{h}}(n+1) &= \hat{\mathbf{h}}(n) + \mu \mathbf{X}_s(n) \left[\mathbf{X}_s^T(n) \mathbf{X}_s(n) \right]^{-1} \mathbf{e}(n) \end{aligned} .$$

Die Länge der Vektoren $\mathbf{x}_z(n)$, $\hat{\mathbf{x}}_z(n)$ und $\mathbf{e}(n)$ bestimmt die AP-Ordnung Q , die der Dimension des affinen Raumes entspricht. Beim NLMS-Algorithmus besitzt der Vektor des Substitutionssignals \mathbf{x}_s die Länge der Filterordnung P . Die AP führt zu einer Erweiterung des Vektors in eine $P \times Q$ Matrix \mathbf{X}_s der Form:

$$\begin{aligned} \mathbf{X}_s(n) &= \begin{bmatrix} \mathbf{x}_s(n) & \mathbf{x}_s(n-1) & \dots & \mathbf{x}_s(n-Q+1) \end{bmatrix} \text{ und} \\ \mathbf{x}_s(n) &= \begin{bmatrix} x_s(n) & x_s(n-1) & \dots & x_s(n-P+1) \end{bmatrix}^T. \end{aligned}$$

Die Gültigkeit von x_z (bzw. \mathbf{x}_z) wird über das Statusbit indiziert, das zusammen mit den Signaldaten blockweise übertragen wird. Signalisiert das Statusbit einen Ausfall, so werden die letzten gültigen Filterkoeffizienten "eingefroren" und an das Rekonstruktionsfilter weitergeleitet. Gleichzeitig liegt an dessen Eingang das Substitutionssignal, das durch die Filterung mit $\hat{\mathbf{h}}$ eine Nachbildung des Zielsignals \tilde{x}_z liefert,

$$\tilde{x}_z(l) = \mathbf{x}_s^T(l) \hat{\mathbf{h}}_n, \quad 1 \leq l \leq L,$$

wobei l den Index des ausgefallenen Signalsamples kennzeichnet und die gesamte Länge des Ausfalls mit L definiert ist. Das Subscript n bei den Filterkoeffizienten weist darauf hin, dass diese zum Zeitpunkt n berechnet und anschließend festgehalten werden.

In Abb. 4-2 ist der Übergang von Zielsignal auf dessen Schätzung während eines Ausfalls schematisch als harter Schalter dargestellt. In der Praxis erfolgt dies durch eine Überblendung der beiden Signale. Bei den einkanaligen Anordnungen in Kapitel 2 wurde auf

die Problematik einer Überblendung in einem latenzzeitkritischen System eingegangen. Die Schwierigkeit liegt darin, dass die Information über die Gültigkeit eines Datenblocks erst im entsprechenden Block mitgeliefert wird. Wäre eine "Früherkennung" eines Ausfalls möglich, so könnte der Übergang zwischen Zielsignal und dessen Schätzung durch einfaches Aus- bzw. Einblenden realisiert werden.

Ein mehrkanaliges Verschleierungsverfahren, das in Kombination mit einer speziellen Signalkodierung als Quelle-Filter-Modell verwendet wird, wird von Karadimou et. al. [Karadimou 2006] vorgeschlagen. Dieser Ansatz ist jedoch nicht mehr von der vorgeschalteten Kodierer-Dekodierer-Einheit entkoppelt. Die zu übertragenden Kanäle setzen sich aus einem Audioreferenzkanal und Parameterkanälen zusammen, letztere beinhalten die LSF-Vektoren (*line spectral frequencies*) aller Signale [Bäckström 2004]. Um eine für alle Signale akzeptierbare Ausfallsverschleierung zu gewährleisten, muss der Referenzkanal Informationen aller Signale enthalten. In vielen praktischen Mehrkanalanordnungen kann diese Voraussetzung allerdings nicht erfüllt werden. Weiters wird vorausgesetzt, dass der Referenzkanal jedenfalls fehlerfrei übertragen und nie durch einen Ausfall gestört wird. Die Einschränkung widerspricht der Anforderung an das zu entwickelnde M-DROCON-System, das in ein beliebiges mehrkanaliges Szenario inkludiert werden soll. Rechenkomplexität und Zeitverzögerung werden in [Karadimou 2006] nicht sehr kritisch betrachtet, was einerseits durch die Oktavband-Filterbank mittels Daubechies-Wavelet-Filter [Daubechies 1992] und andererseits durch die Implementierung eines *Gaussian mixture model* [Deller 2000] zur Schätzung der LSF-Vektoren bei Signalausfall offensichtlich wird. Da die gesamte Signalverarbeitung direkt im Signalweg des gestörten Kanals stattfindet, ist eine Verschleierungstechnik mit ultrakurzer Latenzzeit nicht realisierbar.

Die Verwertung bzw. Adaptierung der genannten Verfahren in das dieser Arbeit zugrunde liegende Gesamtsystem erweist sich als schwierig und führt allgemein zu einer nicht zufriedenstellenden Qualität der verschleierten Signale. Ein Grund dafür liegt einerseits in den speziellen Kodierer-Dekodierer-Modulen, auf denen die Methoden von Sinha oder Karadimou beruhen. Andererseits spielt – vor allem bei der Verwendung der adaptiven Filter

– die konkrete Signalstruktur der Mehrkanalanordnung eine wesentliche Rolle. Diese ist nichtstationär sowohl was die Signale der Einzelquellen als auch die Konfiguration (Anzahl und Stärke) der gerade aktiven Quellen (= Instrumente) angeht. Die teils großen räumlichen Distanzen, die etwa bei der Surroundmikrofonierung eines Orchesters auftreten, sind mit adaptiven Filtern nur bedingt handhabbar. Als Folge der geometrischen Gegebenheiten ist die Kreuzkorrelation zwischen Ziel- und Substitutionskanal gegebenenfalls gering. Daher wären lange Adaptionszeitkonstanten notwendig. Dieser Umstand steht in direktem Konflikt zu der nichtstationären Eigenschaft der Eingangssignale, die eine starke Schwankung der Kreuzkorrelation nach sich zieht und somit eine schnelle Adjustierung der Filter erfordern würde.

Der Versuch, mithilfe einiger informeller Tests einen allgemein passenden Wert (d. h. eine Kompromisslösung) für die Lernrate zu finden, lieferte für keinen der verwendeten adaptiven Algorithmen eine annehmbare Qualität der nachgebildeten Signale.

Im nächsten Abschnitt wird ein Verfahren vorgestellt, welches das Problem der unzureichenden Korrelationsschätzung versucht zu umgehen und auf der Verarbeitung der Betragsspektren beruht. Dies führt zu einem stabileren Filter und weiters zu einer deutlich verbesserten Qualität des geschätzten Zielsignals.

4.2. Spektrales Verschleierungsfilter

Kennzeichnend für die neue Verschleierungsmethode ist die Berechnung eines linearphasiges FIR-Filters im Frequenzbereich, das auf die Nachbildung der Phasenbeziehung zwischen den Signalen verzichtet und ausschließlich auf den Zusammenhang zwischen den spektralen Energieverteilungen ausgerichtet ist. Dieser Ansatz verbessert die Stabilität des Filters und führt zu einer deutlich verbesserten Qualität des geschätzten Zielsignals .

Als Voraussetzung für eine adäquate Schätzung der Filterkoeffizienten müssen die jeweils betrachteten Signale fehlerfrei vorliegen. Bei Auftreten eines Ausfalls im Zielsignal werden die zuletzt ermittelten Koeffizienten in den Zeitbereich transformiert und damit die entsprechenden Samples des intakten Substitutionssignals gefiltert. Die dem linearphasigen Filter inhärente Zeitverzögerung wirkt sich unmittelbar auf die Schätzung des Zielsignals aus und muss dementsprechend berücksichtigt werden. Im Unterschied dazu verursacht weder die Frequenztransformation der Signale noch die Rücktransformation der Koeffizienten eine zusätzliche Latenz im aktiven Signalpfad, da die Koeffizientenberechnung von der Filterung selbst entkoppelt ist (*offline* Verarbeitung). Abb. 4-3 zeigt einen Ausschnitt aus den aktiven Signalpfaden von Ziel- und Substitutionssignal, der – abhängig vom Statusbit des Zielsignals¹⁸ – in drei Teile unterteilt ist:

(A) Sektion unmittelbar vor Ausfall von x_Z , ($do_Z = false$):

Teil (A) entspricht dem Pufferinhalt der Signale, der über eine Transformation $T\{ \}$ in den Frequenzbereich gebracht und dort zur Schätzung des Verschleierungsfilters H verwendet wird. Diese Schritte werden laufend ausgeführt, solange die Bedingung $do_Z = false$ erfüllt ist.

(B) Ausfall von x_Z , ($do_Z = true$):

Die inverse Transformation $T^{-1}\{ \}$ von H wird einmalig zu Beginn des Ausfalls von x_Z berechnet. Anschließend wird blockweise die Nachbildung des Zielsignals \hat{x}_Z erzeugt.

(C) Bereich nach Ausfall von x_Z , ($do_Z = false$):

Nach Beendigung des Ausfalls wird das wieder intakte Zielsignal fortgesetzt. Ab diesem Zeitpunkt wird die grün/strichliert umrandete Einheit wieder aktiviert (ist in der Grafik nicht mehr explizit spezifiziert).

Die Darstellung beinhaltet die wesentlichen Elemente des Basisalgorithmus, der im nächsten Abschnitt im Detail vorgestellt wird.

¹⁸ Das Statusbit des Substitutionssignals ist nicht angeführt, es wird hier ein ausfallsfreies x_S angenommen.

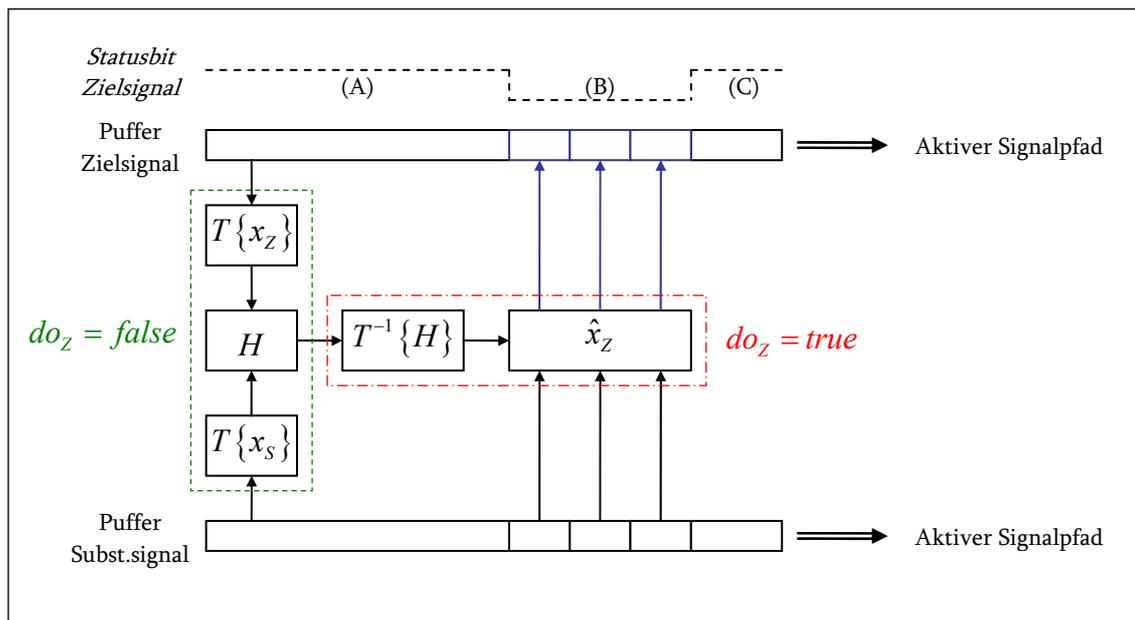


Abb. 4-3: Ausschnitt aus Signalpfad,
Elemente des Basisalgorithmus

In Abschnitt 4.2.2 wird eine Erweiterungsstufe erläutert, deren Ziel die Berücksichtigung der gegebenenfalls aufgrund der räumlichen Anordnung der Mikrofone entstehenden Laufzeitdifferenz zwischen den Signalen in Form eines Zeitgliedes τ_2 ist. Dessen Berücksichtigung im aktiven Signalpfad von x_s dient zur Synchronisierung von Zielsignal und dessen Schätzung und somit zur Verbesserung des Signalübergangs.

4.2.1. Basisalgorithmus

Die Berechnung des Verschleierungsfilters im Frequenzbereich resultiert in einer blockweisen Verarbeitung mit Blocklängen, die einem Vielfachen der Paketgröße der Übertragung entspricht. Damit ist auch für diesen Algorithmus eine Pufferung der vergangenen Daten aller Kanäle notwendig.

Zur Ermittlung des Amplitudengangs eines Filters dienen nur die Einhüllenden der Betragsspektren von Ziel- und Substitutionssignal, nicht die schmalbandigen Spektren. Die Aufgabe des Filters besteht darin, die Klangcharakteristik (Formanten, Klangfärbung, etc.) des Zielsignals zu erfassen und auf dessen Schätzung zu übertragen, die während des Ausfalls durch die Filterung des Substitutionssignals gewonnen wird. Klangverfärbungen, die als

Folge der Ausfallsverschleierung auftreten können, wirken sich nachteilig auf die Qualität des Audiosignals aus. Deshalb ist das Filter, das die Einhüllenden der Signale in Beziehung stellt, für die Beibehaltung der spektralen Eigenschaften im geschätzten Signal verantwortlich.

Die Ermittlung der Einhüllenden kann auf unterschiedliche Art implementiert werden, wobei die Berechnung der diskreten Fouriertransformation (DFT) mit sehr kurzer Blocklänge (d. h. geringer Frequenzauflösung) die direkteste und recheneffizienteste Methode darstellt. Alternativ kann eine Kurzzeit-DFT mit einer größeren Blocklänge (typischerweise 512 bis 2048 Punkten bei $F_s=44.1\text{kHz}$) implementiert werden. Die Einhüllende des Betragsspektrums errechnet sich über die Detektion der lokalen Maxima und anschließender Interpolation zwischen diesen Stellen. Ein Vorteil dieser Variante liegt darin, dass zur Ermittlung der Kohärenzfunktion bzw. des Kreuzleistungsdichtespektrums, die für die Berechnung des Zeitgliedes (Abschnitt 4.2.2) und für die Auswahl des Substitutionskanals (Abschnitt 4.2.3) benötigt werden, DFT's in derselben Größenordnung eingesetzt werden und infolgedessen eine gemeinsame Transformation ausreichend ist.

Darüber hinaus sind auch komplexere Techniken wie beispielsweise das reellwertige Cepstrum (mit Fensterung im Cepstralbereich und anschließender Rücktransformation in den Frequenzbereich und Entlogarithmierung des modifizierten Spektrums [Deller 2000]), die Wavelet-Transformation (ebenfalls inkl. Betragsbildung [Daubechies 1992]) oder die Implementierung einer Gammaton-Filterbank, deren Effektivwert je Kanal verwendet wird, [Iriño 2001] denkbar. Diese Methoden werden hier nicht weiter berücksichtigt. Primär wird die Einhüllende aus der Kurzzeit-DFT ermittelt. Abb. 4-4 zeigt einen Ausschnitt des schmalbandigen Spektrums und im Vergleich dazu dessen Einhüllende eines beliebigen Audiosignals.

In der nachfolgenden Beschreibung wird vorausgesetzt, dass das Substitutionssignal während der gesamten Berechnung fehlerfrei verfügbar ist. Die Bestimmungsmethoden von x_s werden in Abschnitt 4.2.3 vorgestellt.

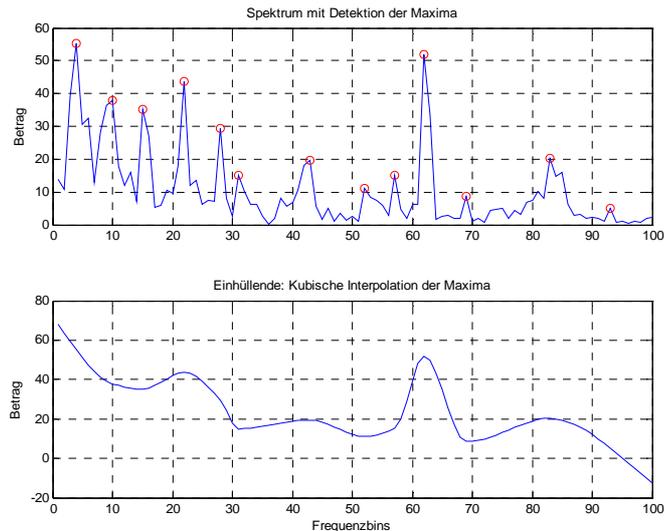


Abb. 4-4: Berechnung der Einhüllenden des Betragsspektrums:
kubische Interpolation der Maxima

Schätzung des spektralen Verschleierungsfilters

Das spektrale Verschleierungsfiler zieht im Gegensatz zu den vorhin vorgestellten, zeitbereichsbasierten Verfahren nicht die Minimierung der Differenz zwischen Zielsignal und dessen Schätzung zur Aktualisierung der Filterkoeffizienten heran. Vielmehr werden die Einhüllenden der Betragsspektren von Ziel- und Substitutionssignal in Beziehung gesetzt.

Im Frequenzbereich errechnet sich das Fehlersignal aus:

$$E(k) = \left| S_z(k) \right| - \left| \hat{S}_z(k) \right| = \left| S_z(k) \right| - H(k) \left| S_s(k) \right|,$$

wobei $\left| S_z(k) \right|$, $\left| \hat{S}_z(k) \right|$, $\left| S_s(k) \right|$ jeweils die Einhüllende des Betragsspektrums des Zielsignals, dessen Schätzung sowie des Substitutionssignals bezeichnen und $H(k)$ das spektrale Filter darstellt. Über die Minimierung des quadratischen Fehlers ergibt sich im einfachsten Fall folgende Formulierung für $H(k)$:

$$H(k) = \frac{\left| S_z(k) \right|}{\left| S_s(k) \right|}. \quad (4.1)$$

Die Berechnung ist einmalig bei Detektion des Ausfalls im Zielkanal mithilfe der im Puffer gespeicherten Signaldaten durchzuführen. In der Verarbeitung von nichtstationären Eingangssignalen erweist es sich jedoch als vorteilhaft, deren Verlauf ebenfalls miteinzubeziehen, was durch die zeitliche Mittelung der Betragsspektren erreicht wird. Ein einfaches Mittelungsverfahren bietet der gewichtete gleitende Mittelwert (*moving average*),

$$\overline{|S_Z(k, m)|} = \sum_{p=0}^P b_p(k) |S_Z(k, m-p)|,$$

dessen Koeffizienten bzw. Ordnung durch $b_p(k)$ bzw. P gegeben sind und m den Blockindex (des Übertragungsformats) darstellt. Aufgrund der Implementierung auf spektraler Ebene sind die Koeffizienten frequenzabhängig.

Für das spektrale Verschleierungsfilter wird die Verwendung einer exponentiellen Glättung mit frequenzunabhängiger Glättungskonstante $\alpha, [0 < \alpha \leq 1]$, bevorzugt, die für Ziel- und Substitutionssignal gleich gewählt wird:

$$\overline{|S_Z(k, m)|} = \alpha |S_Z(k, m)| + (1-\alpha) \overline{|S_Z(k, m-1)|},$$

$$\overline{|S_S(k, m)|} = \alpha |S_S(k, m)| + (1-\alpha) \overline{|S_S(k, m-1)|}.$$

Als Verallgemeinerung der Mittelung wird eine nichtlineare Verzerrung der Betragsspektren implementiert, die speziell die zeitliche Varianz der Signalenergie berücksichtigt und folglich in beiden Signalen getrennt einzustellen ist. Eine Potenzfunktion mit den (frei wählbaren) Exponenten γ bzw. δ bietet den Vorteil, im zeitvarianten Verlauf jedes Frequenzbereichs¹⁹ Zeiträume mit hoher bzw. geringer Signalenergie in ihrem Einfluss auf das Ergebnis der zeitlichen Mittelung in diesem Frequenzbereich verschieden zu gewichten. So entsprechen Exponenten γ und δ größer 1 einer Expansion, d. h. die im zeitlichen Verlauf auftretenden Signalspitzen beherrschen das Ergebnis der zeitlichen Mittelung, wohingegen Exponenten kleiner 1 eine Kompression darstellen, die Zeiträume mit geringer Signalenergie stärker berücksichtigt. Die optimale Wahl der Exponentenwerte hängt vom zu erwartenden Klangmaterial ab, allgemein wird ein Wertebereich von $0.5 \leq \gamma, \delta \leq 2$

¹⁹ In der vorliegenden Realisierung entspricht ein Frequenzbereich (-band) exakt einem Frequenzbin, da die Einhüllende des Spektrums durch Interpolation erzeugt wird. Bei anderen Anwendungen wie z. B. der Gammaton-Filterbank erscheint der Ausdruck "Bereich" bzw. "Band" angemessener.

vorgeschlagen. Damit werden die exponentiell verzerrten, gemittelten Betragsspektren wie folgt formuliert:

$$\overline{|S_Z(k, m)|} = \left\{ \alpha |S_Z(k, m)|^\gamma + (1 - \alpha) \overline{|S_Z(k, m-1)|}^\gamma \right\}^{\frac{1}{\gamma}}, \quad (4.2)$$

$$\overline{|S_S(k, m)|} = \left\{ \alpha |S_S(k, m)|^\delta + (1 - \alpha) \overline{|S_S(k, m-1)|}^\delta \right\}^{\frac{1}{\delta}}. \quad (4.3)$$

Als alternative Methoden zur nichtlinearen Verzerrung können z. B. auch Logarithmus- oder Exponentialfunktionen verwendet werden. Die Berechnung des Filters $H(k)$ aus Gl. (4.1) wird nun durch das Verhältnis von Gl. (4.2) zu Gl. (4.3) realisiert. Die Beschränkung von $H(k)$ wird durch eine Regularisierung vorgenommen, die verhindern soll, dass bei zu schwacher Signalleistung von $\overline{|S_S|}$ die Filterverstärkung zu stark ansteigt und dadurch Hintergrundrauschen hörbar oder das System perceptiv instabil wird. Liegen beispielsweise die Spitzen der spektralen Einhüllenden von $\overline{|S_Z|}$ und $\overline{|S_S|}$ in einem zeitlichen Block nicht exakt im selben Frequenzband, wird $H(k)$ in jenen Bändern, in denen $\overline{|S_Z|}$ ein Maximum und $\overline{|S_S|}$ ein Minimum aufweist, übermäßig ansteigen. Um diesem Problem entgegenzuwirken, wird eine Beschränkung für $H(k)$ durch den frequenzabhängigen Regularisierungsparameter $\beta(k)$ eingeführt, mit

$$H(k) = \frac{\overline{|S_Z(k)|} \overline{|S_S(k)|}}{\overline{|S_S(k)|}^2 + \beta(k)}. \quad (4.4)$$

Durch positiv reellwertiges $\beta(k)$ wird die Filterverstärkung auch bei kleinem Betrag von $\overline{|S_S|}$ nicht zu stark ansteigen und dadurch unerwünschte Signalspitzen verhindert. Diese Vorgangsweise ist durch die Tatsache motiviert, dass in mehrkanaligen Aufnahmen eine fehlerhafte Überhöhung eines Kanals oder Frequenzbereichs auffälliger und damit störender ist als eine fehlerhafte Absenkung, die meist durch die Signale anderer Kanäle maskiert wird.

Die optimalen Werte für $\beta(k)$ sind von der zu erwartenden Signalstatistik abhängig, wobei eine Berechnung auf Basis einer Schätzung der Grundgeräuschleistung in jedem Frequenzband zielführend ist. Die Grundgeräuschleistung $P_g(k)$ kann durch eine zeitlich gemittelte Minimumstatistik geschätzt werden. Für die Berechnung des Regularisierungsparameters $\beta(k)$ wird eine Größe vorgeschlagen, die proportional zum Effektivwert des Grundgeräusches ist: $\beta(k) = c \cdot [P_g(k)]^{\frac{1}{2}}$, wobei c typischerweise Werte zwischen 1 und 5 einnimmt.

Als Alternative kann eine spezielle Formulierung für $H(k)$ herangezogen werden, in der die zeitliche Mittelung anstatt auf die spektralen Einhüllenden direkt auf die Filterkoeffizienten angewandt wird:

$$\overline{H(m,k)} = \left\{ \alpha \left[\frac{|S_z(m,k)| |S_s(m,k)|}{|S_s(m,k)|^2 + \beta(k)} \right]^\gamma + (1-\alpha) \overline{H(m-1,k)}^\gamma \right\}^{\frac{1}{\gamma}}. \quad (4.5)$$

Der Vorteil dieser Variante liegt in der Verringerung des Rechenaufwands. Allerdings besitzt die Formulierung den Nachteil, dass die nichtlineare Verzerrung für die Einhüllenden nicht getrennt vorgenommen werden kann, sondern nur als verallgemeinerte Mittelung der Filterkoeffizienten.

Schätzung des Zielsignals

Die Filterung des Substitutionssignals wird im Zeitbereich implementiert. Grund dafür ist das für den zugrunde liegenden Anwendungsfall charakteristische Echtzeitsystem, das extrem strenge Anforderungen an die Latenzzeit der gesamten Übertragungskette stellt. Im M-DROCON-Modul wird infolgedessen auf Puffereinheiten in den aktiven Signalpfaden verzichtet. Für eine frequenzbereichsbasierte Filterung wäre jedoch die Verfügbarkeit eines Blocks des Substitutionssignals der Länge N (DFT-Länge) während des Ausfalls notwendig, was ausschließlich durch Pufferung erreicht werden kann. Zwar impliziert die linearphasige

Filterung eine Verzögerung τ_1 der halben Filterlänge, durch die aufgrund des breitbandigen Filterverlaufs und einer geeigneten Fensterung eher kurze Impulsantwort kann diese de facto vernachlässigt werden.²⁰ Die inverse Transformation des Verschleierungsfilters ist lediglich zu Beginn eines Ausfalls zu berechnen, danach ist die laufende Schätzung solange deaktiviert, bis die Übertragung des Zielsignals wieder korrekt funktioniert.

Abhängig von der verwendeten Formulierung Gl. (4.4) bzw. Gl. (4.5) erhält man

$$h_w(n) = w(n)T^{-1}\{H(k)\} \text{ bzw. } \overline{h_w(n)} = w(n)T^{-1}\{\overline{H(k)}\},$$

wobei $w(n)$ die Fensterfunktion repräsentiert.

Zur sampleweisen Schätzung des Zielsignals \hat{x}_z ist ein entsprechender Vektor der Länge N des Substitutionssignals \mathbf{x}_s erforderlich,

$$\hat{x}_z(n) = \mathbf{h}_w^T \mathbf{x}_s(n) \text{ bzw. } \hat{x}_z(n) = \overline{\mathbf{h}_w}^T \mathbf{x}_s(n). \quad (4.6)$$

Gl. (4.6) wird in jedem Block berechnet, in dem ein Ausfall vorliegt. Im Anschluss wird das geschätzte Zielsignal noch so lange weitergeführt, bis eine Überblendung in das originale Zielsignal vollzogen ist. Die Länge des Abschnitts ist so zu wählen, dass die Phasenkontinuität beim Signalwechsel gewährleistet wird, d. h. keine hörbaren Artefakte auftreten. Danach werden wiederum ausschließlich die Blöcke des Zielkanals an den Ausgang der Verschleierungseinheit gelegt, sofern keine neuerliche Störung detektiert wird. Der kritischere Übergang zu Beginn des Ausfalls (vgl. Kapitel 2, Abb. 2-2) muss – wie bei den einkanaligen Systemen – mit der Extrapolationsmethode (vgl. Kapitel 2, Gl. (2-15)) realisiert werden. Es ist daher zusätzlich ein lineares Prädiktionsfehlerfilter notwendig.

4.2.2. Zeitglied zwischen Ziel- und Substitutionssignal

Die Erweiterung des Basisalgorithmus zielt darauf ab, die aufgrund der räumlichen Anordnung der Mikrofone entstehenden Laufzeitdifferenzen zwischen den Kanälen festzustellen und nach Möglichkeit dazu zu nutzen, um die Signalphasen von Zielsignal und

²⁰ In der Erweiterung des Basisalgorithmus wird die Verzögerung des linearphasigen Filters mitberücksichtigt.

dessen Schätzung am Beginn des Ausfalls soweit zu synchronisieren, dass auf eine Implementierung der Extrapolation verzichtet werden kann. Gleichzeitig wird auch das durch die linearphasige Filterung verursachte Zeitglied τ_1 miteinbezogen. Da der bisher beschriebene Algorithmus alle für eine Verschleierung erforderlichen Signalmanipulationen beinhaltet, ist der Einsatz der folgenden Berechnungsschritte optional.

Die Ermittlung der "dominanten" Zeitverzögerung τ_2 zwischen Ziel- und Substitutionssignal erfolgt mithilfe der verallgemeinerten Kreuzkorrelation (*generalised cross correlation*, GCC), die über die inverse Fouriertransformation aus dem verallgemeinerten Kreuzleistungsdichtespektrum (*generalised cross power spectral density*, GXPSD) abgeleitet wird, [Carter 1987], [Omologo 1997]. Zur Schätzung des GXPSD dienen die komplexen Spektren der zugrunde liegenden Signale:

$$\Phi_{G, ZS}(k) = G(k) X_Z^*(k) X_S(k).$$

Die Zeitverzögerung τ_2 wird über das Maximum der GCC bestimmt, wobei dessen Detektion umso einfacher wird, je mehr sich die Spitzen der GCC einer Deltafunktion annähern. Das Vorfilter $G(k)$ beeinflusst direkt die Form der GCC und dient somit bei geeigneter Wahl zur Verbesserung der τ_2 -Schätzung. Carter schlägt in [Carter 1987] unter anderem das Phasentransformationsfilter (*phase transform*, PHAT) vor:

$$G_{PHAT}(k) = \frac{1}{|X_Z^*(k) X_S(k)|}.$$

Damit ergibt sich für das GXPSD:

$$\Phi_{G, ZS}(k) = \frac{X_Z^*(k) X_S(k)}{|X_Z^*(k) X_S(k)|} = \frac{\Phi_{ZS}(k)}{|X_Z^*(k) X_S(k)|}. \quad (4.7)$$

Die Verwendung eines alternativen Vorfilters führt zur Definition der Kohärenzfunktion, welche z. B. in Beamforming-Anwendungen häufig zur Schätzung von Zeitgliedern eingesetzt wird:

$$\Gamma_{ZS}(k) = \frac{\Phi_{ZS}(k)}{\sqrt{\Phi_{ZZ}(k) \Phi_{SS}(k)}}. \quad (4.8)$$

Hierin berücksichtigt das Vorfilter die Autoleistungsdichtespektren des Zielsignals $\Phi_{ZZ}(k)$ und des entsprechenden Substitutionssignals $\Phi_{SS}(k)$.

Die Transformation der Signale in den Frequenzbereich erfolgt standardmäßig mittels DFT, deren Blocklänge den zu erwarteten Zeitverzögerungen angepasst werden muss. Aus diesem Grund wird zur Berechnung der Einhüllenden der Betragsspektren $|S_Z|$ und $|S_S|$ (bei der Schätzung des Verschleierungsfilters) ebenso die Kurzzeit-DFT herangezogen und damit kein zusätzlicher Rechenaufwand verursacht. Um die Varianz der Kurzzeitschätzung von $\Phi_{G,ZS}(k)$ bzw. $\Gamma_{ZS}(k)$ zu reduzieren, ist eine (zeitliche) Mittelung der Funktionen erforderlich, die in Form einer exponentiellen Glättung mit den Glättungskonstanten λ bzw. ν , $0 < \lambda, \nu \leq 1$, realisiert wird:

$$\overline{\Phi_{G,ZS}(m,k)} = \lambda \frac{\Phi_{ZS}(m,k)}{|X_Z^*(m,k)X_S(m,k)|} + (1-\lambda)\overline{\Phi_{G,ZS}(m-1,k)}, \quad (4.9)$$

$$\overline{\Gamma_{ZS}(m,k)} = \nu \frac{\Phi_{ZS}(m,k)}{\sqrt{\Phi_{ZZ}(m,k)\Phi_{SS}(m,k)}} + (1-\nu)\overline{\Gamma_{ZS}(m-1,k)}. \quad (4.10)$$

Zusätzlich zum Frequenzindex k ist in beiden Formulierungen der Blockindex m angeführt, um die zeitliche Abhängigkeit des Mittelungsverfahrens zu verdeutlichen. Aus den gemittelten Funktionen ergibt sich schließlich die verallgemeinerte Kreuzkorrelation:

$$\phi_{ZS}(n) = \text{IDFT} \left\{ \overline{\Phi_{G,ZS}(k)} \right\} \quad \text{bzw.} \quad \phi_{ZS}(n) = \text{IDFT} \left\{ \overline{\Gamma_{ZS}(k)} \right\},$$

wobei der Index n in diesem Fall den Versatz zwischen Ziel- und Substitutionssignal darstellt. Je nach Anordnung der Mikrofone der beiden Kanäle kann n sowohl positive als auch negative Werte annehmen, es gilt: $\left(-\frac{N}{2} + 1 \leq n \leq \frac{N}{2} \right)$ und $N =$ Blocklänge der DFT. Für eine kausale Zeitverzögerung τ_2 darf jedoch nur jene Hälfte von $\phi_{ZS}(n)$ in Betracht gezogen werden, in der n negativ ist, d. h. das Signal im Zielkanal muss bezogen auf das im Substitutionskanal verzögert sein. Diese Konstellation ermöglicht die Verwendung von gepufferten Blöcken von x_S zur Schätzung des Zielsignals, wohingegen im Basisalgorithmus

jeweils nur der aktuelle Block verwertet wird. Positives n würde hingegen bedeuten, dass x_s x_z "nacheilt" und folglich zukünftige Blöcke von x_s erforderlich wären, welche einzig durch die Implementierung eines Puffers in den aktiven Signalweg verfügbar gemacht werden können. Wo dies nicht möglich ist, wird ein positives τ_2 als ungültig gekennzeichnet und nicht weiter betrachtet. Der Übergang von Zielsignal auf dessen Schätzung kann somit nicht durch eine Synchronisation verbessert werden. Die Zeitverzögerung τ_2 lässt sich direkt aus dem Index des Maximums von $\phi_{zs}(n)$ ableiten, $\tau_2 = \arg \max_n \{ \phi_{zs}(n) \}$ und $\phi_{zs}(\tau_2) = \max \{ \phi_{zs}(n) \}$.

Die Bedingung für die Gültigkeit der Zeitverzögerung lautet demnach: $\tau_2 \leq 0$. Das effektive Zeitglied, das in den Signalpfad von x_s vor dessen Filterung mit h_w inkludiert wird, errechnet sich letztlich aus $\Delta\tau = |\tau_2| - \tau_1$ und bewirkt eine Optimierung des Verschleierungsverfahrens.

4.2.3. Auswahl des Substitutionssignals

Die optimale Auswahl eines Substitutionskanals ist vom statistischen Zusammenhang zwischen Substitutions- und Zielsignal abhängig, der im Allgemeinen aufgrund der gewählten Mehrkanalanordnung vom Anwender grob abgeschätzt und folglich manuell eine fixe Zuordnung der Kanäle eingestellt werden kann. Diese Methode ist einfach und erfordert keinen zusätzlichen Rechenaufwand, sie besitzt allerdings zwei wesentliche Nachteile. Fällt während eines Ausfalls im Zielsignal auch das Substitutionssignal aus, so stehen zur Schätzung des Zielsignals keine gültigen Daten zur Verfügung. Um die Ausfallssicherheit dennoch zu gewährleisten, muss somit die Möglichkeit eines Austausches von x_s gegeben sein. Zum Zweiten ist die Anwendung auf ein statisches System beschränkt, d. h. es können weder bewegte Quellen noch bewegte Mikrofone berücksichtigt werden. Darüber hinaus kann durch die Kombination mehrerer Kanäle ein besser geeignetes Substitutionssignal generiert und die Qualität der Verschleierungseinheit verbessert werden. Eine

entsprechende Zusammenstellung von x_s bedarf einiger Berechnungen und kann nicht im Vorhinein fix festgelegt werden.

Die Signalähnlichkeit zwischen den einzelnen Kanälen kann grundsätzlich über die Schätzung der Kreuzkorrelation bzw. des Kreuzleistungsdichtespektrums bestimmt werden. Im vorgeschlagenen Algorithmus wird die Kohärenzfunktion, die unter anderem zur Bestimmung der Zeitverzögerung τ_2 dient, herangezogen. Zunächst werden Kanalpaare definiert, die aus dem Zielkanal und einem Substitutionskandidaten zusammengesetzt sind. Die maximale Anzahl der Alternativkanäle beträgt $K-1$, wenn K die Anzahl der Kanäle der gesamten Anordnung repräsentiert. In der Praxis kann die Menge der Alternativkanäle vom Anwender (z. B. Toningenieur) eingeschränkt werden, da die Charakteristiken der einzelnen Kanäle, d. h. deren Signalähnlichkeit, aufgrund der bekannten Anordnung grob eingeschätzt werden kann. Von der in Gl. (4.10) definierten zeitlich gemittelten Kohärenzfunktion wird für jedes Kanalpaar j der Betrag $\left| \overline{\Gamma_{zs,j}}(k) \right|$ betrachtet. Als Auswahlparameter für x_s dient der über die Frequenz gemittelte Betrag von $\left| \overline{\Gamma_{zs,j}}(k) \right|$, in nachstehender Formulierung repräsentiert N die Blocklänge der DFT:

$$\chi(j) = \frac{1}{N} \sum_{k=0}^{N-1} \left| \overline{\Gamma_{zs,j}}(k) \right|.$$

Letztlich kennzeichnet der Index J des Maximalwerts der Funktion $\chi(j)$ jenen Alternativkanal, dessen Signalinhalt die beste statistische Äquivalenz zum Zielsignal aufweist:

$$x_s = x_J, \quad J = \arg \max_j \{ \chi(j) \}. \quad (4.11)$$

Die Verwendung eines einzigen Substitutionskanals ist in der Regel ausreichend, dennoch zeigt sich die "Reservierung einer Reserve" vorteilhaft für den Fall, dass die Übertragung des gewählten ebenso ausfällt. Durch Kontrollieren der Statusbits von x_s und den entsprechenden Reservekanälen kann bei Bedarf direkt auf ein fehlerfreies Signal

umgeschaltet werden. Solange die Kohärenzfunktionen der betrachteten Kanäle keine zu großen Abweichungen voneinander aufweisen, wirkt sich die Auswechslung von x_s nicht nachteilig auf die Qualität des geschätzten Zielsignals aus.

Zur Vermeidung der Umschaltung zwischen verschiedenen Substitutionskanälen wird die Überlagerung aller Kanäle verwendet. Mithilfe der Funktion $\chi(j)$ lässt sich eine passende Gewichtung realisieren:

$$x_s = \frac{\sum_j \{\chi(j)x_j\}}{\sum_j \chi(j)}.$$

Mit dieser Formulierung wird der Einfluss eines Ausfalls in einem der Signale x_j reduziert. Einmal mehr empfiehlt sich, im Vorhinein eine Einschränkung der potenziellen Kanäle x_j durch den Anwender vorzunehmen. Anstelle einer manuellen Vorauswahl wird eine automatisierte vorgeschlagen, die einerseits mithilfe eines Schwellwertes Θ für die Maxima von $\chi(j)$ realisiert werden kann:

$$\tilde{J} = \left\{ j \mid (1 \leq j \leq K-1) \wedge (\chi(j) > \Theta) \right\}. \quad (4.12)$$

Andererseits kann zur Begrenzung eine Maximalanzahl M_j von beispielsweise zwei bis fünf Kanälen zugelassen werden:

$$\tilde{J} = \left\{ j_i \mid (1 \leq j_i \leq K-1) \wedge (1 \leq i \leq M_j) \wedge \left[\chi(j_i) > \chi(l), \forall l \in \{1, \dots, K-1\} \setminus \{j_1, \dots, j_{M_j}\} \right] \right\}. \quad (4.13)$$

Die gemeinsame Verwendung beider Varianten der Kanalreduktion Gl. (4.12) und Gl. (4.13) ist darüber hinaus möglich.

Bei der Zusammensetzung von x_s müssen letztlich neben der Überprüfung der Signalinhalte noch zwei Kriterien berücksichtigt werden: (a) das Statusbit der gewählten x_j und (b) deren Zeitverzögerungen $\tau_{2,j}$. Jene Kanäle x_j , bei denen ein gültiges $\tau_{2,j}$ detektiert wird, werden mit $\Delta\tau_j = \tau_{2,j} - \tau_1$ synchronisiert überlagert, der Rest wird vernachlässigt. Als Endformel für die sampleweise Berechnung des Substitutionssignals ergibt sich:

$$x_s(n) = \frac{\sum_j \{\chi(\tilde{J})x_j(n - \Delta\tau_j)\}}{\sum_j \chi(\tilde{J})} \quad \forall \{do(\tilde{J}) = false\}.$$

Einen schematischen Aufbau des Basisalgorithmus samt Erweiterungsstufe zeigt Abb. 4-5. Der Einfachheit halber sind parallele Signale (DFT-Blöcke) bzw. daraus ermittelte (spektrale) Abbildungen in einer (durchgezogenen) Linie zusammengefasst, deren Anzahl mit K oder $K-1$ gegeben ist. Die strichlierten Verbindungen kennzeichnen die Übertragung bzw. Eingabe von Parametern. Im Block "Selektor" erfolgt die erste Auswahl der Substitutionskanäle aufgrund der Kohärenzfunktion. Das zweite Auswahlkriterium liefert die Zeitverzögerung τ_2 . Die Statusbits der Kanäle sind nicht explizit dargestellt, deren Verifizierung findet in den relevanten signalverarbeitenden Blocks statt. Somit kann in dieser Darstellung auf die explizite Kennzeichnung des Zielsignals verzichtet werden.

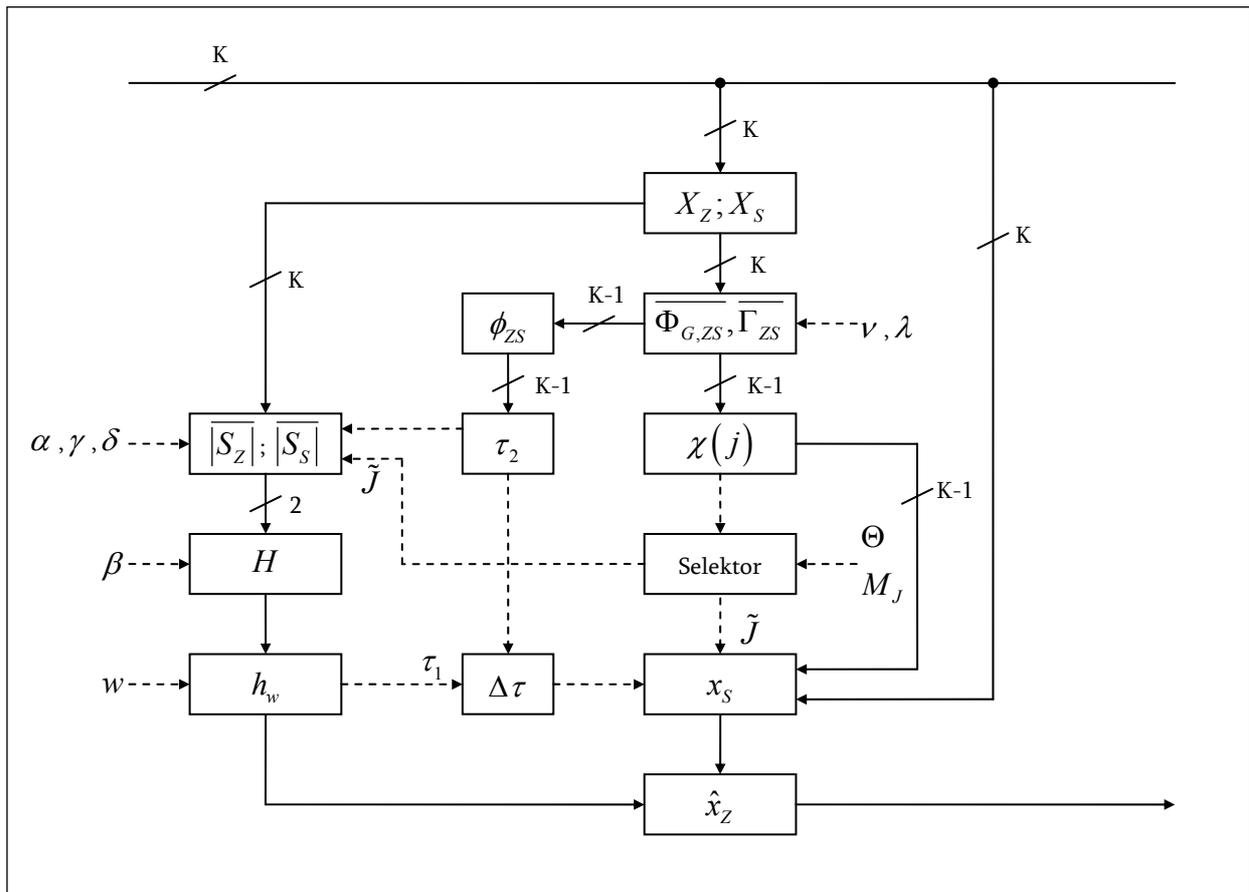


Abb. 4-5: Flussdiagramm des gesamten Algorithmus

4.3. Zusammenfassung

Die Interpretation der Signalausfallsverschleierung in Form eines MIMO-Systems stellt eine Erweiterung der einkanaligen Methoden dar. Als Hauptanwendungsgebiet wird aktuell die Übermittlung von Audiodaten in Netzwerken gesehen, in der die Behandlung von Übertragungsfehlern vorwiegend gemeinsam auf Sender- und Empfängerseite implementiert wird [Perkins 1998], [Wah 2000a]. Für das M-DROCON-Modul wird hingegen ein rein empfängerseitig basiertes Verfahren entwickelt, das wiederum das Optimierungsziel einer minimalen Verarbeitungslatenz verfolgt. Der Übergang von *intra-channel-* auf *inter-channel-* Strategien führt zu einem vergrößerten Datenraum, der zur Schätzung des Zielsignals in

Betracht gezogen wird, siehe Abb. 4-6. Da neben den zeitgleich vorhandenen Blöcken der parallelen Kanäle (bezogen auf den ausgefallenen Block eines Kanals) auch deren Signalvergangenheit gepuffert und damit verfügbar ist, kann in der Extremsituation eines Totalausfalls dennoch ein mehrkanaliger Verschleierungsansatz realisiert werden.

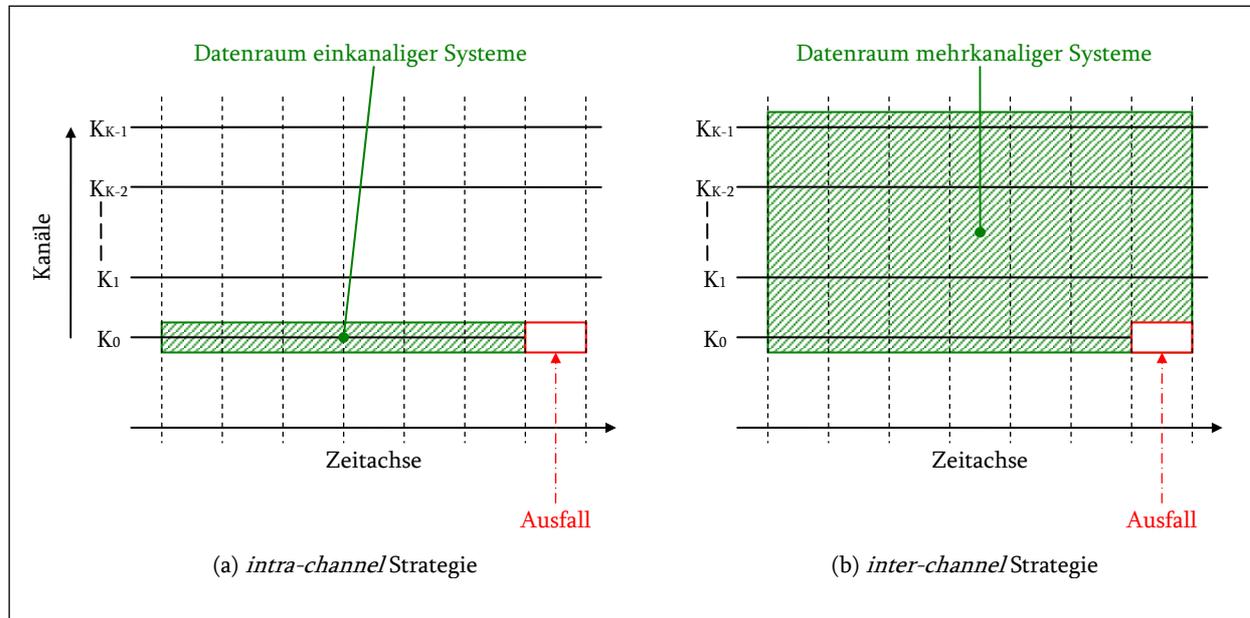


Abb. 4-6: Verfügbare Datenräume zur Verschleierung
(a) *intra-channel* Strategie, (b) *inter-channel* Strategie

Die Ausnutzung der Signalkorrelation in mehrkanaligen Systemen findet in weiten Bereichen des Arrayprocessing Anwendung, z. B. zur Ausrichtung von Mikrofonarrays [Brandstein 2001], Trennung von Nutz- und Störsignal [Van Veen 1988], etc. Die Geometrie der eingesetzten Mikrofonarrays ist meist so abgestimmt, dass die Abstände zwischen den einzelnen Sensoren deutlich geringer sind als in der vorliegenden Betrachtung und daher stabile Korrelationen zwischen den einzelnen Kanälen zu erwarten sind. Deshalb führt Signalverarbeitung mithilfe von adaptiven Filtern zu leistungsfähigen Systemen. Fällt die stabile Korrelation zwischen den Kanälen weg, weil der Abstand zwischen den Mikrofonen für die meist hallige Aufnahmeumgebung zu groß ist und mehrere unabhängige Quellen in sehr unterschiedlichen Distanzen zu den Mikrofonen aktiv sind (typisch für kombinierte Haupt-Stütz-Mikrofonierung bei größeren Klangkörpern), so sind adaptive Filter, die auf stabilen Phasenbeziehungen zwischen Substitutions- und Zielkanal basieren, ungeeignet.

Für diesen Fall wird ein spektrales Verschleierungsfilter vorgeschlagen, das aus dem Verhältnis der spektralen Betragsverläufe von Ziel- und Substitutionssignal geschätzt wird. Die Vernachlässigung der Phasenbeziehungen wird entweder durch ein geeignetes Überblenden zu Beginn und am Ende der Signalerzeugung oder durch die Verarbeitung der "dominanten" Laufzeitdifferenzen zwischen den betrachteten Signalen in einer separaten Einheit kompensiert.

Das in diesem Kapitel entwickelte Verfahren kann allgemein zur Fehlerbehandlung in beliebigen Mehrkanalsystemen verwendet werden und ist durch eine internationale Patentschrift geschützt. Die Offenlegung der Erfindung befindet sich in Anhang B.

Kapitel 5

Resümee und Ausblick

Im Verlauf der vorliegenden Arbeit wurden Verschleierungsalgorithmen untersucht, die in ein Echtzeitsystem integrierbar sind, das die extreme Anforderung der "Null-Latenz" (d. h. keine Pufferung im aktiven Signalpfad) erfüllt. Dabei erwies es sich als vorteilhaft, eine getrennte Behandlung von ein- und mehrkanaligen Systemen vorzunehmen.

Bei den Einkanalssystemen wurde von bekannten Methoden ausgegangen, die wirkungsvoll in vergleichbaren Anwendungen eingesetzt werden. Die Erweiterung der Verfahren bestand vor allem darin, vorhandene Verzögerungsglieder aus dem aktiven Signalpfad zu eliminieren. Dies wurde durch eine Kombination aus Extrapolation und Substitution erreicht, wobei eine recheneffiziente Implementierung mithilfe einer in Echtzeit realisierbaren Interpolation entwickelt wurde. Die Verarbeitung wurde im Zeitbereich vorgenommen, eine Aufbereitung der Eingangssignale (z. B. eine Transformation in den Spektralbereich oder eine Subbandzerlegung) wurde nicht berücksichtigt.

Die Verschleierungsmethoden in Mehrkanalsystemen sind derzeit noch wenig untersucht, die vorhandenen Ansätze sind jeweils für ein spezielles Anwendungsziel entwickelt worden. In Kapitel 4 wurden diese Verfahren zusammengefasst und verglichen. Es wurde untersucht, inwieweit deren Einsatz bei mehrkanaligen Mikrofonierungstechniken (Haupt- und Stützmikrofonie) möglich ist. Aufgrund der aufgezeigten Probleme wurde ein neuer Ansatz zur mehrkanaligen Verschleierung eingeführt.

Entsprechend der Unterteilung der gesamten Arbeit ist auch der Ausblick zweigeteilt und liefert für jeden Bereich einen separaten Ansatz zur Weiterentwicklung der gewonnenen Erkenntnisse.

Einkanalige Systeme

Die in Kapitel 2 beschriebenen Methoden zur Signalfortführung – Extrapolation, Substitution, Interpolation – erfordern die Bestimmung der Autokorrelation des intakten Signals vor einem Ausfall. Für deren korrekte Berechnung ist eine Fensterung notwendig, $w(n)=1 \quad \forall \quad 0 \leq n \leq N-1$ und $w(n)=0 \quad \text{sonst}$. In den vorgeschlagenen Algorithmen wird auf diese Fensterung allerdings verzichtet, was zur Verwendung der Autokovarianzfolge führt, die in der Verarbeitung von nichtstationären Signalen gegenüber der Autokorrelation Vorteile mit sich bringt, [Makhoul 1975].

Mithilfe des Blockschaltbildes in Abb. 2-1, das die grundsätzliche Funktionsweise der Ausfallsverschleierung darstellt, lässt sich eine Einteilung der gesamten Berechnungsschritte in Abhängigkeit des Signalzustandes festlegen:

Signalzustand	Berechnungen
Signal intakt, kein Ausfall	keine
Beginn des Ausfalls	Autokovarianzfolge Prädiktionskoeffizienten erster Block der Verschleierung
während des Ausfalls	fortlaufende Verschleierung
nach dem Ausfall	fortlaufende Verschleierung für Überblendung

Tab. 5-1: Prozessorauslastung in Abhängigkeit des Signalzustandes

Aus Tab. 5-1 ist ersichtlich, dass Autokovarianzfolge und Prädiktionskoeffizienten jeweils nur im ersten Block eines Ausfalls bestimmt werden müssen, während bei intakt arbeitender Übertragungsstrecke keine Rechenoperationen notwendig sind. Dadurch ergibt sich eine ungleichmäßige Verteilung der Prozessorleistung. Eine verbesserte Auslastung des Prozessors könnte erreicht werden, indem die Berechnung der Autokovarianzfolge

aufgespalten und laufend auf jeden Block aufgeteilt wird. Dazu sollen zwei aufeinanderfolgende Zeitpunkte m und $m+1$ betrachtet werden:

$$\mathbf{r}_m = \begin{bmatrix} r_m(0) \\ r_m(1) \\ \vdots \\ r_m(M-1) \end{bmatrix} = \begin{bmatrix} x(m) & x(m-1) & \dots & x(m-P+1) \\ x(m-1) & x(m-2) & \dots & x(m-P) \\ \vdots & \vdots & \ddots & \vdots \\ x(m-M+1) & x(m-M) & \dots & x(m-M-P+1) \end{bmatrix} \begin{bmatrix} x(m) \\ x(m-1) \\ \vdots \\ x(m-P+1) \end{bmatrix},$$

$$\mathbf{r}_{m+1} = \begin{bmatrix} r_{m+1}(0) \\ r_{m+1}(1) \\ \vdots \\ r_{m+1}(M-1) \end{bmatrix} = \begin{bmatrix} x(m+1) & x(m) & \dots & x(m-P+2) \\ x(m) & x(m-1) & \dots & x(m-P+1) \\ \vdots & \vdots & \ddots & \vdots \\ x(m-M+2) & x(m-M+1) & \dots & x(m-M-P+2) \end{bmatrix} \begin{bmatrix} x(m+1) \\ x(m) \\ \vdots \\ x(m-P+2) \end{bmatrix}. \quad (5.1)$$

Das untere Gleichungssystem kann aus dem oberen folgendermaßen berechnet werden:

$$\mathbf{r}_{m+1} = \mathbf{r}_m - \begin{bmatrix} x(m-P+1) & x(m-P) & \dots & x(m-M-P+1) \end{bmatrix}^T x(m-P+1) + \begin{bmatrix} x(m+1) & x(m) & \dots & x(m-M+2) \end{bmatrix}^T x(m+1) \quad (5.2)$$

Mit dieser Formulierung beträgt der Rechenaufwand je Sample $2M$ Operationen (Multiplikationen/Divisionen), wobei eine weitere Reduktion möglich ist, wenn stattdessen eine Schätzung des Autokovarianzvektors mittels exponentieller Glättung durchgeführt wird:

$$\mathbf{r}_{m+1} = \alpha \mathbf{x}_{m+1} \mathbf{x}_{m+1}^T + (1-\alpha) \mathbf{r}_m, \quad (5.3)$$

wobei α die Zeitkonstante der Mittelung darstellt und

$$\mathbf{x}_{m+1} = \begin{bmatrix} x(m+1) & x(m) & \dots & x(m-M+2) \end{bmatrix}^T.$$

Dem Vorteil der verringerten Anzahl von Operationen steht hier allerdings der Nachteil gegenüber, dass der durch die Näherung entstehende Fehler sich unmittelbar auf die Stabilität der Levinson-Durbin-Rekursion auswirkt, welche die Autokovarianzfolge zur Ermittlung der Prädiktionskoeffizienten benötigt. Eine ausführliche Diskussion bezüglich Fehlerarten sowie Stabilität des rekursiven Algorithmus liefern Papaodysseus et. al. in [Papaodysseus 1993]. Eine erste Untersuchung der modifizierten Implementierung zeigte,

dass eine stabile Koeffizientenschätzung in den meisten Fällen möglich war. Es wurde ein zusätzlicher Vergleich angestellt, bei dem sowohl Gl. (5.2) wie auch Gl. (5.3) nicht mit jedem Sample aktualisiert wurden, sondern jeweils nur zu jedem k -ten Abtastpunkt (was einer Unterabtastung des Eingangssignals entspricht). Dabei war ein schlechteres Stabilitätsverhalten der Levinson-Durbin-Rekursion in Verbindung mit Gl. (5.3) zu erkennen.

Bei der Periodizitätsschätzung kann eine Aufteilung der Berechnung der Differenzfunktion in ähnlicher Weise erfolgen. Andererseits wäre es denkbar, eine einzige Autokovarianzfolge für Koeffizienten- und Periodizitätsschätzung zu realisieren und somit die DF komplett einzusparen. Der Autokovarianzvektor und die entsprechenden Signalvektoren müssten verlängert werden, um eine ausreichende Frequenzauflösung zu erhalten. Dennoch wäre der Rechenaufwand geringer als bei einmaliger Berechnung nach Gl. (5.1).

Mehrkanalige Systeme

Im zweiten Teilbereich der vorliegenden Arbeit wurden Techniken zur Ausfallsverschleierung untersucht, die auf einer gemeinsamen Ausnutzung sämtlicher Kanäle eines Mehrkanalsystems beruhen. Auch hier richtete sich die Vorgehensweise primär auf eine Verarbeitung im Zeitbereich. Aufgrund des speziellen Anwendungsziels, der Gewährleistung einer gegenüber Ausfällen robusten digitalen Drahtlosübertragung von breitbandigen Audiosignalen in Echtzeit, hat sich jedoch gezeigt, dass die Verwendung einer spektralbereichsbasierten Methode wesentliche Vorteile mit sich bringt. Der gegenständlich vorgeschlagene Algorithmus stellt ein Verfahren dar, das generisch formuliert ist und für einzelne Berechnungsschritte mehrere Lösungsvarianten bietet.

Eine zentrale Aufgabe in der Weiterentwicklung des spektralen Verschleierungsfilters besteht darin, diese Vorschläge eingehend zu prüfen, d. h. deren Funktionalität unter Berücksichtigung der vorgegebenen Rahmenbedingungen und der Charakteristik der Eingangssignale zu untersuchen. Die Optimierung des Basisalgorithmus muss darauf

abzielen, geeignete Werte für die Parameter der exponentiellen Glättung α , der nichtlinearen Verzerrung γ , δ sowie der Beschränkung der Filterkoeffizienten $\beta(k)$ zu ermitteln. Vor allem die Adjustierung der Verzerrungsfunktion erscheint für eine bestmögliche Schätzung von $H(k)$ maßgeblich. Der Verbesserungsansatz in der Erweiterungsstufe richtet sich auf die Berechnung des verallgemeinerten Kreuzleistungsdichtespektrums, das als Grundlage für die Schätzung des Zeitgliedes $\Delta\tau$ sowie die Selektion der Kanäle, aus denen das Substitutionssignal zusammengesetzt wird, dient. Aktuell sind dafür die Kohärenzfunktion und das GXPSD mit PHAT-Filter vorgesehen. Die Bestimmung der effektiveren Version kann mithilfe eines Vergleichstests erfolgen.

Weiters ist eine Vereinfachung des Algorithmus durch die Einschränkung der Strategien zur Generierung des Substitutionssignals möglich. Neben den in Abschnitt 4.2.3 angeführten Methoden kann noch ein zusätzlicher Ansatz in Erwägung gezogen werden, der die Kanalauswahl für jedes Frequenzband getrennt beschreibt. Anstelle der frequenzunabhängigen Funktion $\chi(j)$ wird z. B. der Betrag der zeitlich gemittelten Kohärenzfunktion eingesetzt, somit gelten im ersten Selektionsschritt folgende Bedingungen für eine Frequenzkomponente k :

- bei Verwendung eines einzigen Kanals:

$$J(k) = \arg \max_j \left\{ \left| \overline{\Gamma_{ZS,j}(k)} \right| \right\},$$

- bei Verwendung mehrerer überlagerter Kanäle, Schwellwert-Kriterium Θ_k :

$$\tilde{J}(k) = \left\{ j_k \left| \left(1 \leq j_k \leq K-1 \right) \wedge \left(\left| \overline{\Gamma_{ZS,j_k}(k)} \right| > \Theta_k \right) \right. \right\},$$

- bei Verwendung mehrerer überlagerter Kanäle einer maximalen Anzahl $M_{j,k}$ je Frequenzkomponente:

$$\tilde{J}(k) = j_{i,k} \left((1 \leq j_{i,k} \leq K-1) \wedge (1 \leq i_k \leq M_{j,k}) \wedge \left[\left| \overline{\Gamma_{ZS,j_i}(k)} \right| > \left| \overline{\Gamma_{ZS,i_k}(k)} \right|, \forall i_k \in \{1, \dots, K-1\} \setminus \{j_{1,k}, \dots, j_{M_{j,k}}\} \right] \right).$$

Das Substitutionssignal ergibt sich durch eine gewichtete Überlagerung von Bandpasssignalen, die mittels inverser Fouriertransformation aus den entsprechenden Spektren berechnet werden:

$$x_S(n) = \frac{\sum_{\tilde{j}} \left\{ \left| \overline{\Gamma_{ZS,\tilde{j}}(k)} \right| \cdot x_{\tilde{j},k}(n - \Delta\tau_{\tilde{j},k}) \right\}}{\sum_{\tilde{j}} \left| \overline{\Gamma_{ZS,\tilde{j}}(k)} \right|} \quad \text{mit } x_{j,k}(n) = \text{IDFT} \left\{ X_j(k) \right\}.$$

Der aktuelle Stand der Forschung befindet sich noch am Beginn der Entwicklung von mehrkanaligen Ausfallsverschleierungsverfahren. Allerdings zählen deren Einsatzgebiete – mehrkanalige Mikrofonierungssysteme, Datentransfer in Netzwerken, etc. – bereits seit einigen Jahren zu üblichen Anwendungen im Bereich Audio und Multimedia. Der Schritt von einkanaligen zu mehrkanaligen Techniken bedeutet die Einführung einer neuen Generation von Verfahren, deren Potential in der Verbesserung von echtzeitfähigen und im besonderen von latenzzeitkritischen Audioübertragungssystemen in Zukunft noch weit stärker ausgenutzt werden wird.

Anhang A

Perzeptive Evaluierung

A.1. Bewertungsskala

Die Beurteilung der Unterschiede zwischen Referenz- und Vergleichsstimulus (A-B, A-C) erfolgt anhand der in Standard ITU-R BS.1116-1, [ITU-R 1997], [ITU-R 2003], definierten, fünfteiligen Skala.

Verschlechterung	subjektiver Grad
nicht wahrnehmbar	5.0
wahrnehmbar, aber nicht störend	4.0
leicht störend	3.0
störend	2.0
sehr störend	1.0

A.2. Testreihe 1

Subjektiver Hörtest zur Evaluierung der Verschleierungstechniken "Extrapolation" und "Interpolation" hinsichtlich der Qualität der verschleierte Signale.

Unabhängige Faktoren

- F₁ Audiobeispiel: Pop, Jazz, Klassik, Klavier, Geige, Sprache (weiblich).
- F₂ Algorithmus: Extrapolation, Interpolation.
- F₃ Ausfallsszenario: vier zufällig generierte Szenarien mit folgenden Kriterien:
gestörter Signalanteil beträgt 2% der Gesamtlänge eines Beispiels,
Verteilung der Ausfälle:

Länge [ms]	1	2	3	4	5	6	7	8	9	10
Anzahl	2	2	1	2	2	1	2	2	1	2

- F₄ Session: zwei Sessions zur Prüfung der Reliabilität der Subjekte.
- F₅ Subjekt: 40 *non-expert listeners*.

Statistische Auswertung und grafische Darstellungen

Prüfung auf Normalverteilung der Daten: Chi-Quadrat-Test → keine Normalverteilung.

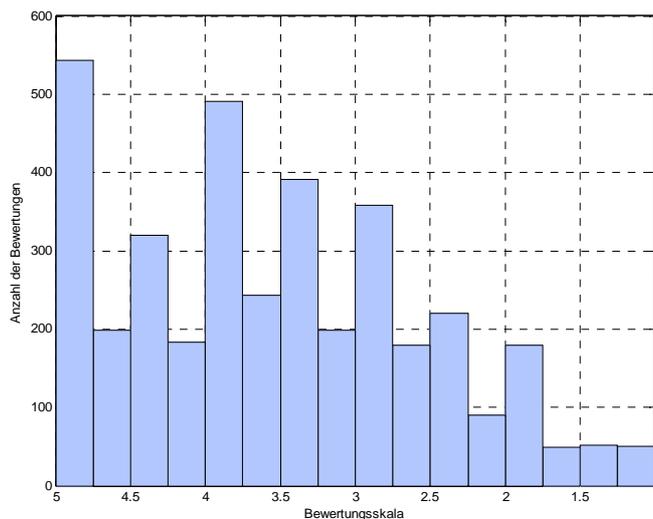


Abb. A-1: Verteilung der subjektiven Bewertung, Testreihe 1

Prüfung auf Signifikanz von F_1 : H-Test → signifikante Unterschiede (5% Level) zwischen den Stichproben.

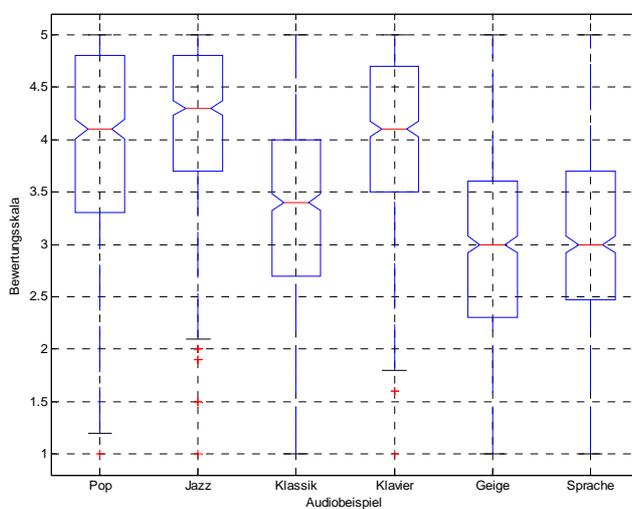


Abb. A-2: Vergleich der Audiobeispiele, Testreihe 1

Prüfung auf Signifikanz von F_2 : U-Test → kein signifikanter Unterschied (5% Level) zwischen den Stichproben.

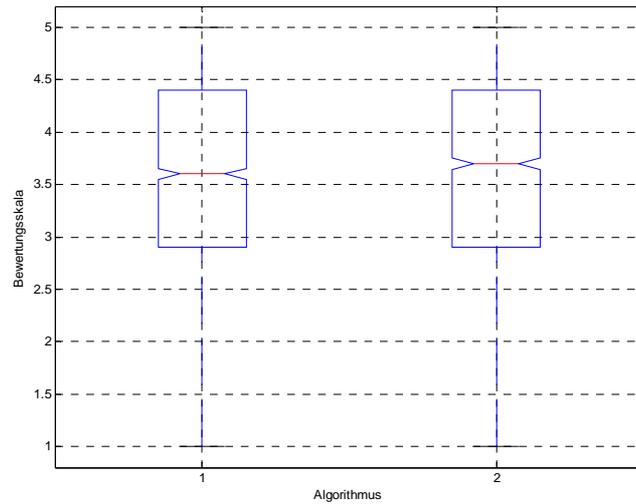


Abb. A-3: Vergleich der Algorithmen, Testreihe 1

Prüfung auf Signifikanz von F_3 : H-Test → signifikanter Unterschied (5% Level) zwischen den Stichproben.

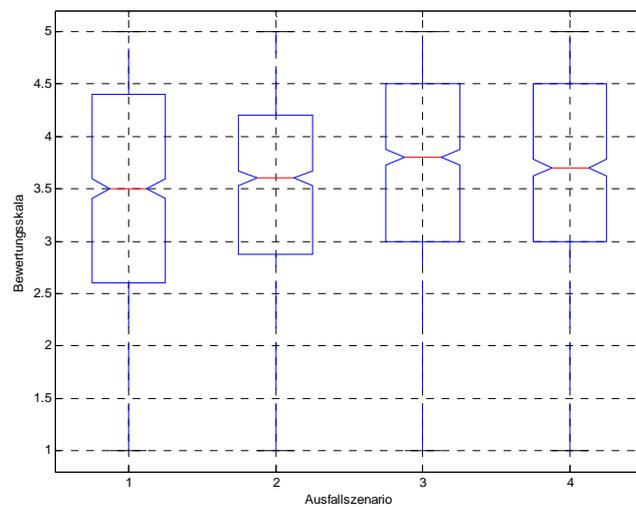


Abb. A-4: Vergleich der Ausfallszenarien, Testreihe 1

Prüfung auf Signifikanz von F_4 : U-Test → signifikanter Unterschied (5% Level) zwischen den Stichproben.

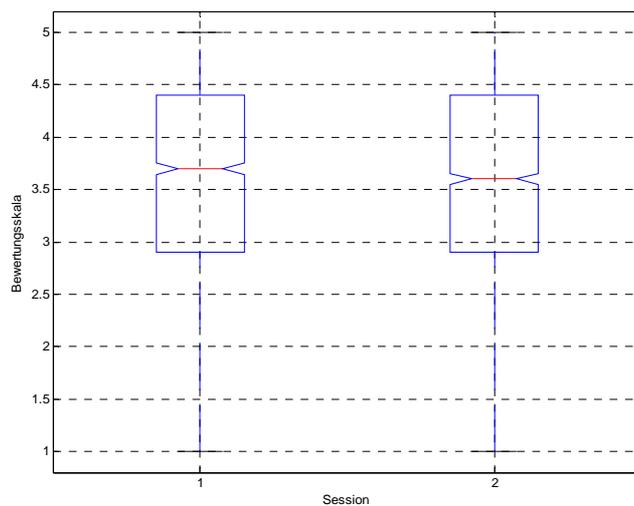


Abb. A-5: Vergleich der Sessions, Testreihe 1

Fehlbewertungen (Klassifizierung, siehe Kapitel 3, Tab. 3-4)

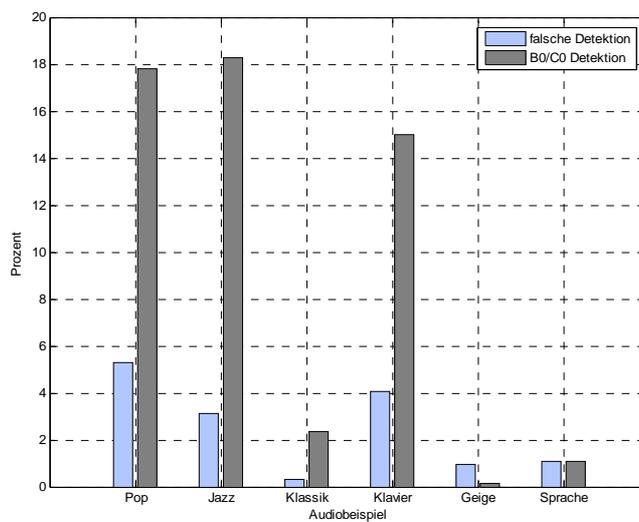


Abb. A-6: Prozentueller Anteil der Fehlbewertungen in Testreihe 1

Kontrolltrials (Definition der Typen von Kontrolltrials, siehe Kapitel 3, Abschnitt 3.2)

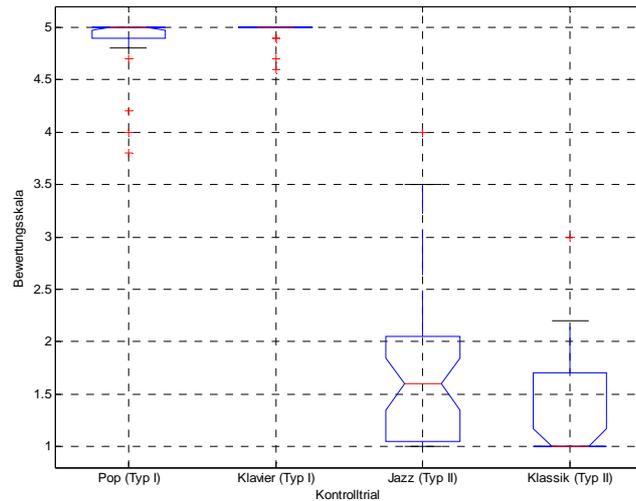


Abb. A-7: Vergleich der Kontrolltrials von Testreihe 1

A.3. Testreihe 2

Subjektiver Hörtest zur Optimierung der Verschleierungsmethode "Extrapolation" – Evaluierung von unterschiedlichen Parametersettings.

Unabhängige Faktoren:

- F₁ Audiobeispiel: Geige, Sprache (weiblich), Saxofon, Gitarre.
- F₂ Parametersetting: drei unterschiedliche Einstellungen der dem Algorithmus inhärenten Parameter.
- F₃ Subjekt: 15 *non-expert listeners*.

Dieser Testreihe liegt ein zufällig generiertes Ausfallszenario mit folgenden Kriterien zugrunde:

- gestörter Signalanteil beträgt 2% der Gesamtlänge eines Beispiels,
- Verteilung der Ausfälle:

Länge [ms]	2	3	4	5	6
Anzahl	9	6	5	4	3

Statistische Auswertung und grafische Darstellungen

Prüfung auf Normalverteilung der Daten: Chi-Quadrat-Test → keine Normalverteilung.

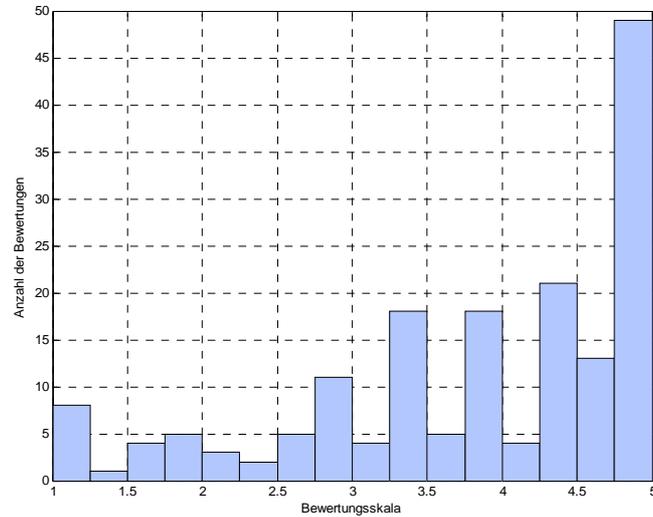


Abb. A-8: Verteilung der subjektiven Bewertung, Testreihe 2

Prüfung auf Signifikanz von F_1 : H-Test → signifikante Unterschiede (5% Level) zwischen den Stichproben.

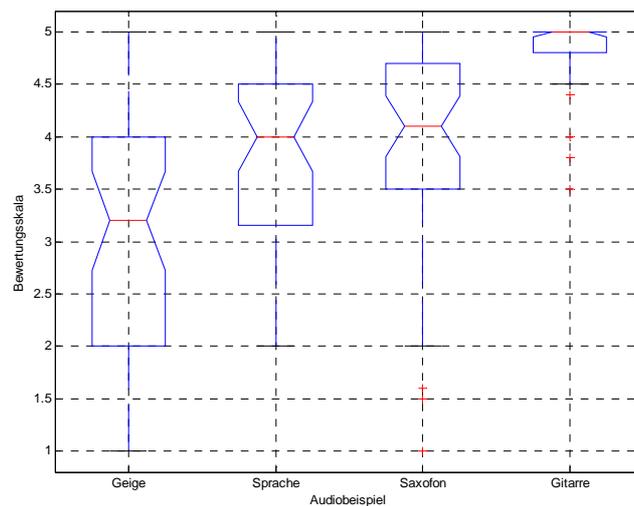


Abb. A-9: Vergleich der Audiobeispiele, Testreihe 2

Prüfung auf Signifikanz von F_2 : H-Test → signifikante Unterschiede (5% Level) zwischen den Stichproben.

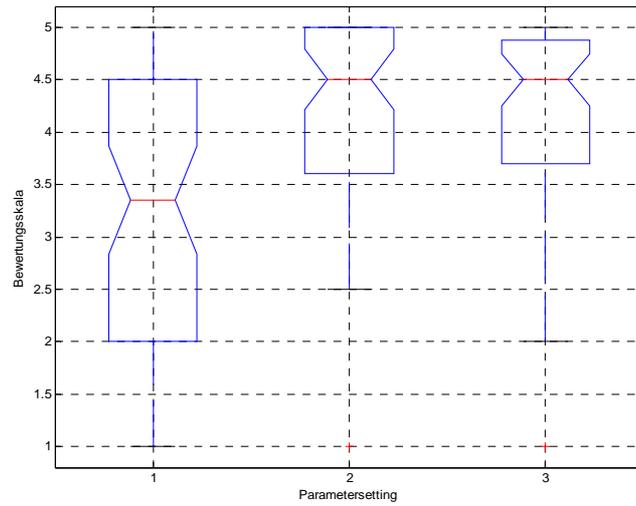


Abb. A-10: Vergleich der Parametersettings, Testreihe 2

Fehlbewertungen (Klassifizierung, siehe Kapitel 3, Tab. 3-4)

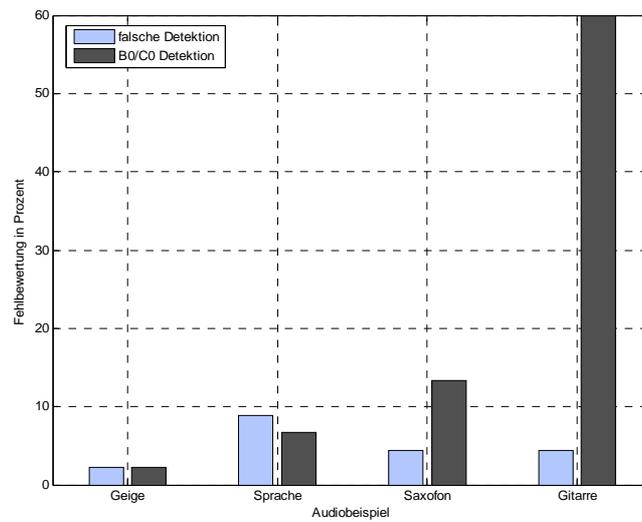


Abb. A-11: Prozentueller Anteil der Fehlbewertungen in Testreihe 2

Kontrolltrials (Definition der Typen von Kontrolltrials, siehe Kapitel 3, Abschnitt 3.2)

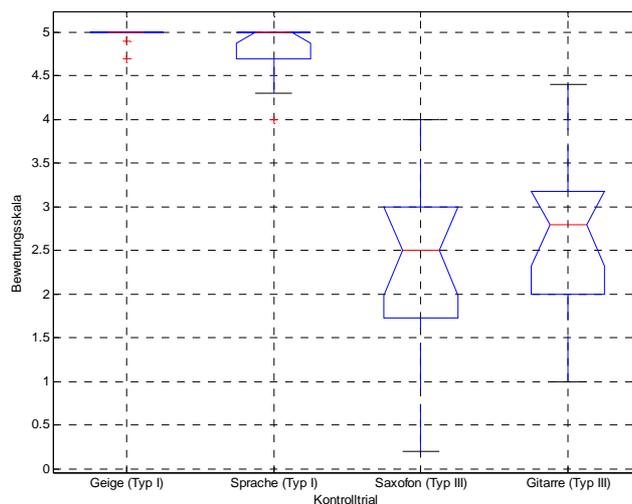


Abb. A-12: Vergleich der Kontrolltrials von Testreihe 2

A.4. Testreihe 3

Subjektiver Hörtest zur Optimierung der Periodizitätsschätzung für die Verschleierungsmethode "Interpolation". Mithilfe dieser Testreihe werden drei unterschiedliche Algorithmen zur Schätzung der Periodizität des Audiosignals evaluiert.

Unabhängige Faktoren

- F₁ Audiobeispiel: Jazz, Klassik, Geige.
- F₂ Algorithmus: Differenzfunktion, "*zero-crossing rate*", AMDF (Beschreibung, siehe Kapitel 2, Abschnitt 2.4.1).
- F₃ Ausfallszenario: zwei zufällig generierte Szenarien mit folgenden Kriterien:
gestörter Signalanteil beträgt 2% der Gesamtlänge eines Beispiels,

Verteilung der Ausfälle:

Länge [ms]	2	3	4	5	6
Anzahl	9	6	5	4	3

- F₄ Subjekt: 21 *non-expert listeners*.

Statistische Auswertung und grafische Darstellungen

Prüfung auf Normalverteilung der Daten: Chi-Quadrat-Test → keine Normalverteilung.

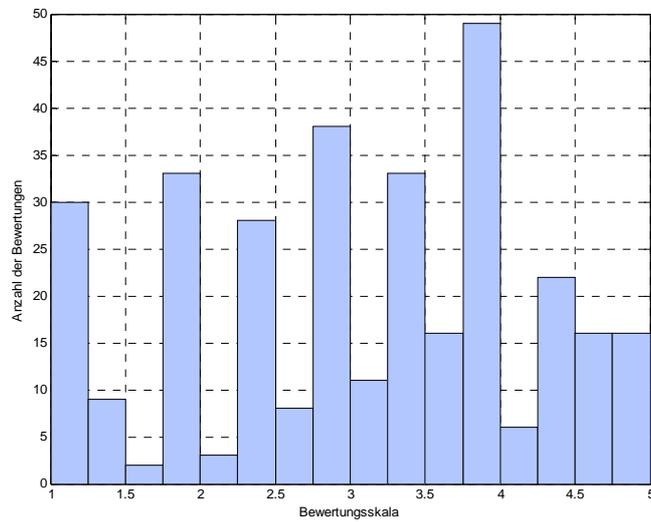


Abb. A-13: Verteilung der subjektiven Bewertung, Testreihe 3

Prüfung auf Signifikanz von F_1 : H-Test → signifikante Unterschiede (5% Level) zwischen den Stichproben.

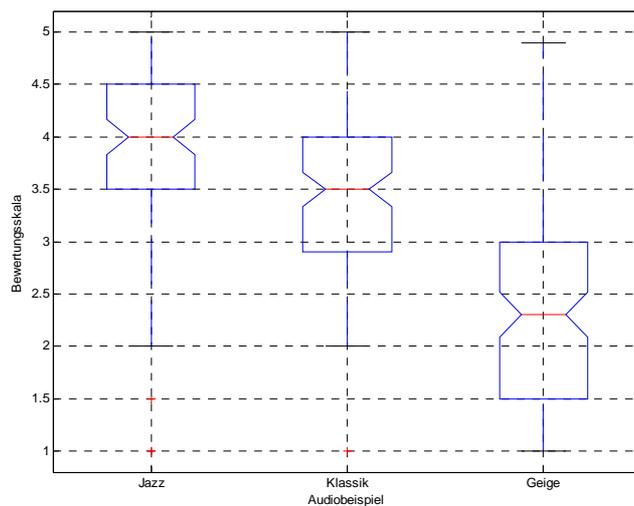


Abb. A-14: Vergleich der Audiobeispiele, Testreihe 3

Prüfung auf Signifikanz von F_2 : H-Test \rightarrow signifikante Unterschiede (5% Level) zwischen den Stichproben.

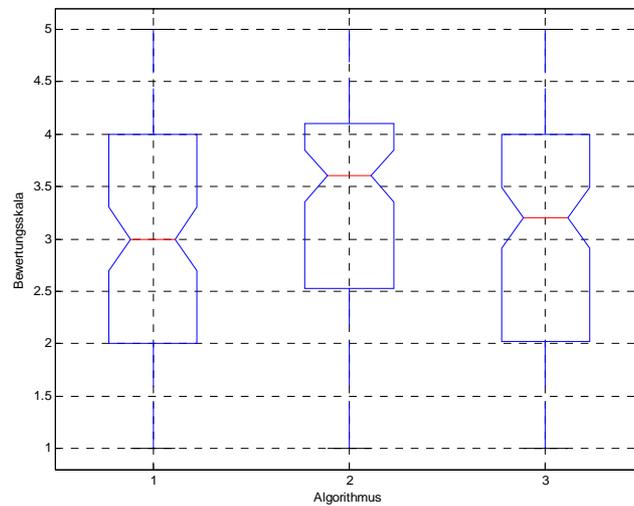


Abb. A-15: Vergleich der Algorithmen, Testreihe 3

Prüfung auf Signifikanz von F_3 : U-Test \rightarrow signifikante Unterschiede (5% Level) zwischen den Stichproben.

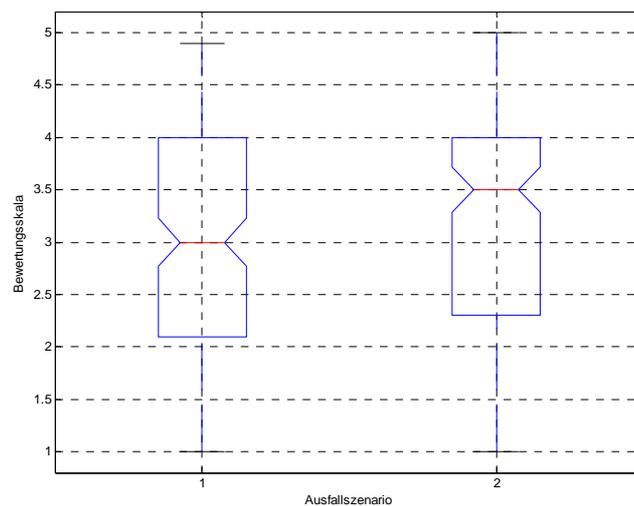


Abb. A-16: Vergleich der Ausfallszenarien, Testreihe 3

Fehlbewertungen (Klassifizierung, siehe Kapitel 3, Tab. 3-4)

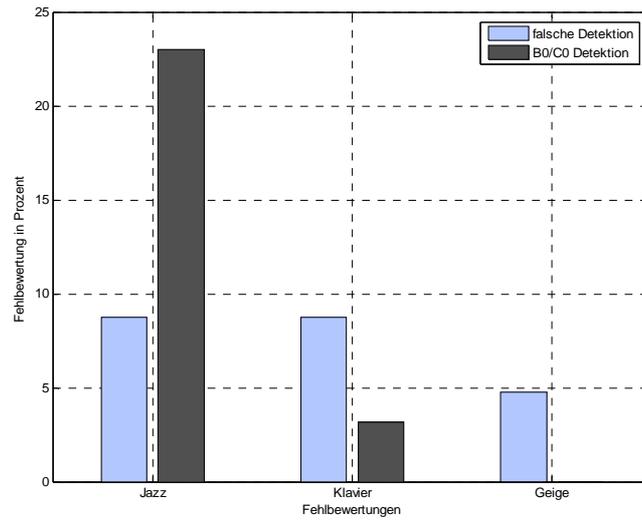


Abb. A-17: Prozentueller Anteil der Fehlbeurteilungen in Testreihe 3

Kontrolltrials (Definition der Typen von Kontrolltrials, siehe Kapitel 3, Abschnitt 3.2)

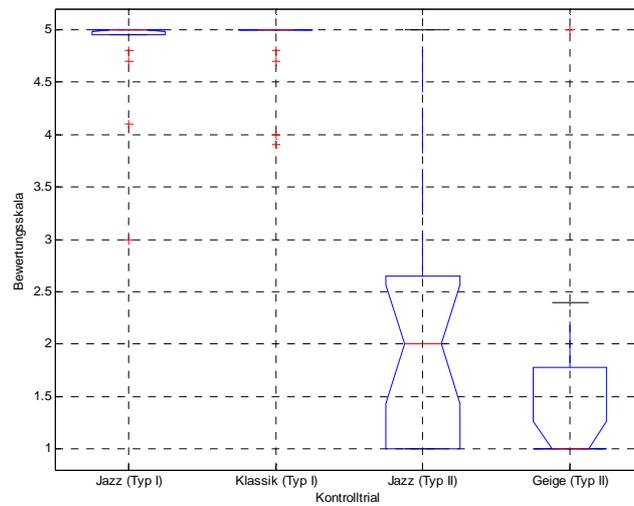


Abb. A-18: Vergleich der Kontrolltrials von Testreihe 3

Anhang B

Patentschrift P43136

Dropout concealment for a multi-channel arrangement

The invention relates to a method for the concealment of dropouts in one or more channels of a multi-channel arrangement comprising at least two channels, wherein a replacement signal is generated in the event of a dropout in one channel with the aid of at least one error-free channel.

The wireless transmission of audio signals has constituted an important area of research since the introduction of the wireless microphone on the market at the beginning of the 1990s. At present, these products are used as standard equipment in the area of stage performances, concerts and live shows. In comparison to analog systems, the use of digital transmission links offers the advantage to transmit metadata in addition to the audio data. This metadata can contain, for example, information about the overall concept of a stage installation. Furthermore, the notion of combining the individual channels and exploiting their interoperability in future systems can be realised by means of digital technologies. Despite, the fast development of the underlying hardware in terms of computing power and storage capacity supports the progress of software implementations.

In general, the method of the wireless transmission of signals is not resistant to influences that can crop up along the transmission link. In case of digital radio links, disturbances directly lead to the loss of data, and hence, to a total signal dropout. The degradation of the signal quality, acoustically perceptible as cracks or clicks, is unacceptable at any rate and must be compensated for using appropriate technologies that are incorporated at the receiver side. Since the concealment unit represents an active element in the signal path, the impact of its inherent processing delay must be taken into consideration.

A general classification of error concealment technologies for audio and video transmissions in real time is offered by Wah B. W., Su X., and Lin D.: "A Survey of Error Concealment Schemes for Real-Time Audio and Video Transmission over the Internet"; Proc. IEEE Int. Symposium on Multimedia Software Engineering, Dec. 2000. Here, the dependence of the

source coding constitutes a fundamental distinguishing characteristic with which a distinction is made between transmitter-controlled and receiver-based technologies. The method according to the invention belongs to the category "receiver-based method", i.e. it works completely decoupled from the transmitter or source coding and is therefore not affected by the additional latency inherent to transmitter-controlled technologies.

The simplest methods for the receiver-based concealment of dropouts are represented by the so-called *intra-channel concealment techniques*, in which each channel of a multi-channel arrangement is treated separately. Standard concealment methods apply substitution and prediction algorithms. The latter are generally comprised by two stages, the analysis unit and the re-synthesis model of the linear prediction error filter. The first stage serves for estimating the filter coefficients and is executed continuously during error-free signal transmission. If a dropout occurs, the lost signal samples are reconstructed by the filtering process. This corresponds to an extrapolation and is suited to the concealment of dropouts of a few milliseconds in general broadband audio signals. In some cases, in which the real-time constraint is not as stringent (for example, the buffering of data is permissible), the extrapolation is transformed into an interpolation and longer dropouts can therefore be handled.

The expansion of one-channel systems to multi-channel systems – the so-called *inter-channel concealment techniques* – leads to the implementation of adaptive filters. Compared to linear prediction algorithms, the estimation of the filter coefficients is not related exclusively to the signal of the respective channel, but rather information from other parallel channels is also used thatfor. The exploitation of the channel cross correlations is deemed to improve the performance of the concealment method. However, the efficiency of the technique is characterised primarily by the convergence behavior of adaptive filters, which mainly depends on the stationarity of the input signals. Since, in general, broadband audio is highly non-stationary, the behaviour of the adaptive filter will be quite poor. One possible implementation of this method is described in US 2005/0182996 A1 (and respective EP

1649452 A1), the entire disclosure of which is incorporated into this specification by virtue of reference.

A common feature of the abovementioned filter techniques denotes the processing in time domain; some algorithms also offer an equivalent description in frequency domain. Yet the aim of the transformation is to increase computing efficiency, whereas the characteristics of the time domain method are retained.

In the following, several concealment methods are described in brief, beginning with single-channel systems:

US 2006/0171373 A1 discloses a single-channel method for the concealment of data losses that makes use of a linear prediction estimate from the intact signal component immediately preceding the dropout. The prediction coefficients obtained by means of a spectral analysis filter are used to estimate a residual signal. A maximum repeatable range is determined for the residual signal over several stages. The spectral analysis of the transmitted signal merely serves for an improved detection of the periodicity, which leads to the classic signal repetition. This period is repeated and the all-pole filter of the linear prediction is applied to it. The residual signal emerges from preceding intact signal components that are filtered inversely with the currently calculated filter coefficients, yielding the estimated replacement signal. All computation required for signal reconstruction is performed in time domain, which is characteristic for the suggested method and results in substantial processing delay. Hence, it is incapable of real-time applications.

DE 19735675 C2 also discloses a single-channel concealment method. The algorithm incorporates a perceptually adapted subband decomposition based on psychoacoustic aspects. The notion of signal reconstruction is to maintain the spectral energy in each subband. If a dropout occurs, an estimation of the signal is obtained by a properly filtered noise signal. Large dropouts yield an unchanged "sound surface". The filter coefficients solely imply the energy information, thus, the preceding time samples are not incorporated.

EP 1 145 227 B1 discloses a single-channel concealment method for the transmission of coded audio signals in the context of the MPEG coding standard. Thus, the transmitted data comprise spectral coefficients rather than time samples. A perceptually adapted subband splitting is employed to the signal section preceding the dropout by combining several MDCT (modified discrete cosine transform) coefficients into one subband. Since a dropout affects certain subbands, these are transformed back into time domain, and a narrow-band signal is predicted there. The estimated narrow-band signal is in turn MDCT-transformed and inserted into the MDCT stream transmitted in MPEG coding.

The article "Packet Loss Concealment for Audio Streaming Based on the GAPES Algorithm" by Ofir et. al. AES 118th Convention, May 28-31, 2005, Barcelona, Spain, describes a single-channel method in the context of the MPEG coding standard and thus, is also MDCT-based. Since the properties of the MDCT prevent an adequate interpolation between successive MDCT blocks, an STFT (short-time Fourier transform) representation is computed directly from the MDCT representation. Interpolation results are obtained in the STFT domain, therefore signal components succeeding the dropout are required, i.e. the method induces additional latency. The interpolation itself is carried out per DFT-bin (discrete Fourier transformation) by use of the GAPES (gapped-data amplitude and phase estimation) algorithm. After the interpolation, the STFT data are transformed back into MDCT data.

The single-channel systems described above essentially depend on past signal components, hence, the estimation of the replacement signal is based on the assumption of long-term stationary input signals. Although those methods that incorporate a spectral analysis apply the filter in the frequency domain, both the comparison with preceding samples and the prediction of the future samples occur exclusively in the time domain.

The article "Packet Loss Concealment for Multichannel Audio Using the Multiband Source/Filter Model" by Karadimou et. al. 40th Annual Asilomar Conf. on Signals, Systems and Computers, Oct. 29 - Nov. 01, 2006, discloses a concealment method that relies on several channels. The transmission format is composed in such a manner that an actual audio

channel is only transmitted in one single, so-called "source channel," whereas LSF (line spectral frequencies) vectors are transmitted in the remaining channels. The LSF vectors represent a (complex valued) spectral interpretation of a time signal and correspond exactly to the linear prediction coefficients. Thus, they contain all of the information on the phase relationships of the spectral envelope. In this method, dropout concealment is constraint to an error-prone "source channel". Dropouts can therefore only be handled in the LSF channels. The estimation of the LSF vectors is done by means of a Gaussian mixture model (GMM). Despite, the method incorporates subband decomposition, per band and channel prediction, and retransformation into linear prediction coefficients with appropriate filtering of the reference residuum. During computation of the replacement signal, i.e. of the LSF vectors, the entire signal information including the phase information is always transmitted. The different LSF vectors of the individual channels contain information about the characteristics of different microphones that are spaced apart from each other, and which simultaneously pick up a sound event, for example a concert. Hence, correlations between the individual LSF vectors are to be expected, and a so-called cross-channel estimation can be employed, i.e. if a dropout occurs in one LSF vector, parallel LSF vectors can be exploited.

For the substitution, a reference channel is established in advance and its LP residuum serves for the signal synthesis of all other channels (not only in the event of dropout but rather during normal operation, too). The fundamental assumption made is that there is a correlation between target and reference channel. However, this assumption is never verified and is definitely not true for many scenarios. The entire processing steps (subband filtering, LP analysis, LSF computation, synthesis filter) of the concealment procedure are implemented in the signal path, resulting in a considerable processing delay that has to be accepted and low latency can not be achieved, respectively. Due to the subband technique, the computational complexity is high (the prediction is performed per subband and channel, and the all-pole filter is implemented in each subband during resynthesis, too).

Another publication that deals with multi-channel concealment is "Loss Concealment for Multi-Channel Streaming Audio" by Sinha et. al. NOSSDAV'03, June 1-3, 2003, Monterey,

California, USA. The particular application of "distributed immersive musical performance" describes the implementation of a collaborative concert of spatially separated musicians by data transfer over the internet. A possibility for signal substitution is suggested therein that is based on the spatial proximity of the loudspeaker positions to each other in the multi-channel setup. In this method, a special type of interleaved packeted transmission is essential for the concealment.

The prior art for multi-channel systems is currently limited to different implementations of adaptive filters in the time domain or on transmitter-side channel interleaving with simple substitution rules as are typical in the upmix/downmix matrixing strategy suggested by Gerzon (M. Gerzon: "Hierarchical System of Surround Sound Transmission for HDTV," AES preprint# 3339, 92nd Convention, March 24-27, 1992, Vienna; and M. Gerzon: "Problems of Upward and Downward Compatibility in Multichannel Stereo Systems," AES preprint# 3404, 93rd Convention, Oct. 1-4, 1992, San Francisco). The efficiency of such technologies is either mostly restricted to its area of application (for example, pre-mixed multi-channel recordings) or is characterised predominantly by the convergence behavior of the adaptive filters, thus is highly variable due to the non-stationary input signals in connection with the dropouts of the target signal.

The aim of the present invention consists in providing a concealment method that uses the intact channels of a multi-channel system to replace the lost signal in such a way that the difference between the original signal and its replacement is rendered inaudible. In addition to the reliability of the transmission, the usability in delay-critical real-time systems constitutes an important criterion, for which reason ultra-low latency techniques are in demand for the processing of signals.

According to the invention, this objective is achieved with a method mentioned at the outset, in that during the error-free signal transmission of the channels a mapping takes place of the transmitted signals into the frequency domain, the absolute value of the frequency spectrum being determined, that spectral filter coefficients are calculated that relate the magnitude spectrum of a channel to the magnitude spectrum of at least one other channel,

and that in the event of the dropout of one channel the replacement signal is generated by computation of filter coefficients prior to the dropout and application of them to a substitution signal which constitutes of at least one error-free channel.

The concealment filter is calculated using the magnitude spectra, thus, without regard to the phase information, providing a more stable filter, and an improved quality of the replacement signal, respectively. A significant advantage compared to single-channel methods currently in use also lies in the utilisation of the interoperability between the individual signals.

As an extension of the basic method, a modified treatment of the phase information is proposed. In so doing, the constancy of the phase transition at the beginning and at the end of the dropout is improved by taking into account the average time delay between target and replacement signal. A time delay between the respective channels, independent of their source direction, emerges according to the spatial arrangement of the multi-channel recording system.

In the following, the invention is described in more detail on the basis of the drawings.

Fig. 1 shows a schematic representation of the transmission chain according to the invention, and

Fig. 2 shows a detailed block diagram of the dropout concealment of the invention for a two-channel system, and

Fig. 3 shows a block diagram of a multi-channel arrangement of, for example eight channels, and

Fig. 4 shows a flowchart of the entire invention, consisting of the estimation of the spectral filter, the determination of the time delay between the channels, as well as the weighted superposition of all channels in order to generate the substitution signal, and

Fig. 5 shows the layout of the device according to the invention for dropout concealment that is to be integrated into each channel of the multi-channel arrangement.

The preferred area of application of the present invention is within the overall system of a multi-channel (optionally wireless) transmission of digital audio data. The entire structure of a transmission chain is depicted in Fig. 1 and typically comprises the following stages for one channel: Signal source 1, e.g. a sensor for recording signals (microphone), analog-digital converter 2 (ADC), optional signal compression and coding on the transmitter side, transmitter 3, transmission channel, receiver 4, concealment module 5. At the output of the concealment module 5, the audio signal is available in digital form – further signal processing units can be connected directly, for example a pre-amp, equalizer, etc.

The proposed concealment method is independent of the transmitter/receiver unit as well as the source coding and acts solely on the receiver side (receiver-based technique). It can therefore be integrated flexibly as an independent module into any transmission path. In some transmission systems (e.g. digital audio streaming), different concealment strategies are implemented simultaneously. While the application shown in Fig. 1 does not provide for any further concealment units, a combination with alternative technologies is possible.

The following application scenarios are provided exemplarily:

- a) In concert events and stage installations, multi-channel arrangements range from stereo recordings to different variations of surround recordings (e.g. OCT Surround, Decca Tree, Hamasaki Square, etc.) potentially supported by different forms of spot microphones. Especially with main microphone setups, the signals of the individual channels are comprised of similar components whose particular composition is often quite non-stationary. For example, a dropout in one main microphone channel can be concealed according to the present invention introducing little or no latency.
- b) Multi-channel audio transmission in studios proceeds at different physical layers (e.g. optical fiber waveguides, AES-EBU, CAT5), and dropouts can occur for various reasons, for example due to loss of synchronization, which must be prevented or concealed especially in critical applications such as, for example, in the transmission operations of a radio station. Here, too, the concealment method according to the invention can be used as a safety unit with a low processing latency.

- c) While audio transmission in the internet is less delay-sensitive than the abovementioned areas, transmission errors occur more frequently, resulting in an increased degradation of the perceptual audio quality. The inventive concealment method offers an improvement of the quality of service.
- d) The method according to the invention can also be used in the framework of a spatially distributed, immersive musical performance, i.e. in the implementation of a collaborative concert of musicians that are separated spatially from each other. In this case, the ultra-low latency processing strategy of proposed algorithm benefits the system's overall delay.

The invention is not restricted to the following embodiment. It is merely intended to explain the inventive principle and to illustrate one possible implementation. In the following, the dropout concealment method is described for *one* channel afflicted with dropouts. If transmission errors occur in more than one channel of the multi-channel arrangement, the system can easily be expanded.

The following terminology is used in the description: The channel afflicted with dropouts is defined as *target channel* or *signal*. The replica (estimation) of this signal that is to be generated during dropout periods is referred to as *replacement signal*. At least one *substitution channel* is required for the computation of the replacement signal.

The proposed algorithm is composed of two parts. Computations of the first part are carried out permanently, whereas the second part is only activated in the case of a dropout in the target channel. During error-free transmission, the coefficients of a linear-phase FIR (finite impulse response) filter of length L_{Filter} are permanently being estimated in the frequency domain. The required information is provided by the optionally non-linearly distorted and optionally time-averaged short-term magnitude spectra of the target and substitution channel. This new type of filter computation disregards any phase information and thus, differs fundamentally from the correlation-dependent adaptive filters.

Selection of the substitution channel or substitution channels

Fig. 2 shows a block diagram of the multi-channel dropout concealment method for a target signal x_z and a substitution signal x_s . The individual steps of the method are each indicated by a box containing a reference symbol and denoted in the subsequent table:

- 6 Transformation into a spectral representation
- 7 Determination of the envelope of the magnitude spectra
- 8 Non-linear distortion (optional)
- 9 Time-averaging (optional)
- 10 Calculation of the filter coefficients
- 11 Time-averaging of the filter coefficients (optional)
- 12 Transformation into the time domain with windowing
- 13 Transformation into the frequency domain (optional)
- 14 Filtering of the substitution signal respectively in time or frequency domain
- 15 Estimation of the complex coherence function or GXPSD
- 16 Time-averaging (optional)
- 17 Estimation of the GCC and maximum detection in the time domain
- 18 Determination of the time delay $\Delta\tau$
- 19 Implementation of the time delay $\Delta\tau$ (optional)

In this example, the transition between target and replacement signal is indicated by a switch 20. A detailed explanation of the individual steps of the method is given in the following description.

The correct selection of a substitution channel depends on the similarity between the substitution and target signal. This correlation can be determined by estimating the cross-correlation or coherence. (See explanations on coherence and on generalized cross-power spectral density (GXPSD) at the end of the specification.) According to the invention, the (GXPSD) is proposed as potential selection strategy. The complex coherence function

$\Gamma_{zS,j}(k)$ is used as particular example in embodiments 1. to 9. (A total of K channels are observed, the channel $x_0(n)$ being designated as the target channel $x_z(n)$):

1. For the target channel $x_z(n)$, the J^{th} channel is defined as a substitution signal by the optionally time-averaged coherence function $\overline{\Gamma_{zS,j}(k)}$ between the channels $x_j(n)$, with $1 \leq j \leq K-1$ and the target channel $x_s(n) = x_j(n)$, whose frequency-averaged value of the complex coherence function, $\chi(j) = \frac{1}{N} \sum_{k=0}^{N-1} \left| \overline{\Gamma_{zS,j}(k)} \right|$, has a maximum value according to: $J = \arg \max_j \chi(j)$.
2. Alternatively, a fixed allocation can be established between the channels in advance if the user (e.g. a sound engineer) knows the characteristics of the individual channels (according to the selected recording method) and hence their joint signal information.
3. Likewise, several channels can be summed to one substitution channel, optionally in a weighted manner. This weighted combination can be set up by the user a priori.
4. In an alternative realisation, the superposition of several channels to one substitution channel is carried out on the basis of broadband coherence ratios to the target

channel by:
$$x_s(n) = \frac{\sum_j \{ \chi(j) \cdot x_j(n - \Delta\tau_j) \}}{\sum_j \chi(j)}, \text{ for all } \{ do(j) = false \}.$$

Herein, $x_s(n)$ denotes the substitution channel composed of the channels $x_j(n - \Delta\tau_j)$, and $\chi(j)$ represents the frequency-averaged coherence function between the target channel $x_z(n)$ and the corresponding channel $x_j(n - \Delta\tau_j)$. The time delay between the selected channel pairs is considered by $\Delta\tau_j$ (c.f. section "Estimation of the time delay between target and substitution channel"). The validity of the potential signals is verified incorporating the status bit $do(j)$.

5. A simplification of 4. is proposed that considers a pre-selected set of channels \tilde{J} rather than all available channels j . The weighted sum is built using $\chi(j) \Big|_{j \in \tilde{J}}$. The

pre-selection is intended to yield channels whose frequency-averaged coherence function exceed a prescribed threshold Θ :

$$\tilde{J} = \left\{ j \mid (1 \leq j \leq K-1) \wedge (\chi(j) > \Theta) \right\}.$$

6. Furthermore, a maximum number of M channels (with preferably $M = 2 \dots 5$) can be established as a criterion, according to:

$$\tilde{J} = \left\{ j_i \mid (1 \leq j_i \leq K-1) \wedge (1 \leq i \leq M) \wedge \left[\chi(j_i) > \chi(l), \forall l \in \{1, \dots, K-1\} \setminus \{j_1, \dots, j_M\} \right] \right\}.$$

7. A joint implementation of both constraints 5. and 6. is also possible:

$$\tilde{J} = \left\{ j_i \mid (1 \leq j_i \leq K-1) \wedge (1 \leq i \leq M) \wedge (\chi(j_i) > \Theta) \wedge \left[\chi(j_i) > \chi(l), \forall l \in \{1, \dots, K-1\} \setminus \{j_1, \dots, j_M\} \right] \right\}.$$

8. Alternatively, the selection can be carried out separately for different frequency bands, i.e. in each band the "optimal" substitution channel is determined on the basis of the coherence function, the respective band pass signals are filtered using the method according to the invention, optionally in a time-delayed manner (c.f. "Estimation of the time delay between target and substitution channel"), superposed and used as a replacement signal. In so doing, the same criteria apply as in 1., 4., 5., 6. and 7., though the frequency-independent function $\left| \overline{\Gamma_{zs,j}(k)} \right|$ must be implemented instead of the frequency-averaged function $\chi(j)$.

9. Several substitution channels can also be selected. In this case, the processing is carried out separately for each channel, i.e. several replacement signals are generated. These are weighted according to their coherence function, combined and inserted into the dropout.

Generally, the functions used in 1. to 9. are time-varying, thus a mathematically exact notation must consider the time dependency by a (block) index m . To simplify the formulations, m has been omitted.

Calculation during error-free transmission

The computation during error-free transmission is performed in frequency domain, thus in a first step an appropriate short-term transformation is necessary, resulting in a block-oriented algorithm that requires a buffering of target and substitution signal. Preferably, the block size should be aligned to the coding format. The estimation of the envelopes of the magnitude spectra of target and substitution signal are used to determine the magnitude response of the concealment filter. The exact narrow-band magnitude spectra of the two signals are not relevant, rather broad-band approximations are sufficient, optionally time-averaged and/or non-linearly distorted by a logarithmic or power function. The estimation of the spectral envelopes can be implemented in various ways. The most efficient possibility concerning computational efficiency is the short-term DFT with short block length, i.e. the spectral resolution is low. A signal block is multiplied by a window function (e.g. Hanning), subjected to the DFT, the magnitude of the short-term DFT is optionally distorted non-linearly and subsequently time-averaged.

Further implementations:

- Wavelet transformation (as is described in Daubechies I.; "Ten Lectures on Wavelets"; Society for Industrial and Applied Mathematics; Capital City Press, ISBN 0-89871-274-2, 1992. The entire disclosure of this printed publication is incorporated into this specification by virtue of reference) with optional subsequent time-averaging of the optionally non-linear distortion of the absolute values of the wavelet transformation.
- Gammatone filter bank (as described in Irino T., Patterson R.D.; "A compressive gammachirp auditory filter for both physiological and psychophysical date"; J. Acoust. Soc. Am., Vol. 109, pp. 2008-2022, 2001. The entire disclosure of this printed publication is incorporated into this specification by virtue of reference) with subsequent formation of the signal envelopes of the individual subbands, optionally followed by a non-linear distortion.
- Linear prediction (as described in Haykin S.; "Adaptive Filter Theory"; Prentice Hall Inc.; Englewood Cliffs; ISBN 0-13-048434-2, 2002. The entire disclosure of this printed publication is incorporated into this specification by virtue of reference) with subsequent sampling of the magnitude of the spectral envelopes of the signal block,

represented by the synthesis filter, optionally followed by a non-linear distortion and, subsequent to this, time-averaging.

- Estimation of the real cepstrum (as described in Deller J.R., Hansen J.H.L., Proakis J.G.; "Discrete-Time Processing of Speech Signals"; IEEE Press; ISBN 0-7803-5386-2, 2000. The entire disclosure of this printed publication is incorporated into this specification by virtue of reference) followed by a retransformation of the cepstrum domain into the frequency domain and taking the antilogarithm, optionally followed by a non-linear distortion of the so obtained envelopes of the magnitude spectra and, subsequent to this, time-averaging.
- Short-term DFT with maximum detection and interpolation: Here, the maxima are detected in the magnitude spectrum of the short-term DFT and the envelope between neighboring maxima are calculated by means of linear or non-linear interpolation, optionally followed by a non-linear distortion of the so obtained envelopes of the magnitude spectra and, subsequent to this, time-averaging.

For the optionally used time-averaging of the envelopes, an exponential smoothing of the optionally non-linearly distorted magnitude spectra can be used, as represented in equations (1) with time constant α for the exponential smoothing. Alternatively, the time-averaging can be formed by a moving average filter. The non-linear distortion can, for example, be carried out by means of a power function with arbitrary exponents which, in addition, can be selected differently for the target and substitution channel, as depicted in equations (1) by the exponents γ and δ . (Alternatively, a logarithmic function can also be used.)

The non-linear distortion offers the advantage of weighting time periods with high or low signal energy differently along the time-varying progression of each frequency component. The different weighting affects the results of time-averaging within the respective frequency component. Accordingly, exponents γ und δ greater than 1 denote an expansion, i.e. peaks along the signal progression dominate the result of the time-averaging, whereas exponents less than 1 signify a compression, i.e. enhance periods with low signal energy. The optimal selection of the exponent values depends on the sound material to be expected.

$$\overline{|S_Z(m)|} = \left\{ \alpha |S_Z|^\gamma + (1-\alpha) \overline{|S_Z(m-1)|}^\gamma \right\}^{\frac{1}{\gamma}}, \quad (1a)$$

$$\overline{|S_S(m)|} = \left\{ \alpha |S_S|^\delta + (1-\alpha) \overline{|S_S(m-1)|}^\delta \right\}^{\frac{1}{\delta}} \quad (1b)$$

where $|S_Z|, |S_S|$: envelopes of the magnitude spectra of target and substitution channel,

$\overline{|S_Z|}, \overline{|S_S|}$: time-averaged versions of $|S_Z|$ and $|S_S|$,

α : time constant of the exponential smoothing, $0 < \alpha \leq 1$,

γ, δ : exponents of the non-linear distortion of $\overline{|S_Z|}$ and $\overline{|S_S|}$, with a preferable value range

of: $0.5 \leq \gamma, \delta \leq 2$,

m : block index.

As an example, equations (1) constitute a special case for the calculation of the spectral envelopes of target and substitution channel with exponential smoothing and arbitrary distortion exponents. In the following, the exponents are set to $\gamma = \delta = 1$ to simplify formulations (i.e. a non-linear distortion is not explicitly indicated). However, the invention comprises the method with any time-averaging methods and any non-linear distortions of the envelopes of the magnitude spectra and hence, any values for the exponents γ and δ . Beyond, the use of the logarithm of the exponential function is enclosed, too. To simplify notation, the block index m is omitted, though all magnitude values such as $\overline{|S_S|}, \overline{|S_Z|}$ or H are considered to be time-variant and therefore a function of block index m .

Calculation of the concealment filter

In standard adaptive systems, concealment filters are calculated by minimizing the mean square error between the target signal and its estimation. The difference signal is given by $e(n) = x_z(n) - \hat{x}_z(n)$. In contrast, the present invention examines the error of the estimated magnitude spectra:

$$E(k) = \overline{|S_Z(k)|} - \overline{|\hat{S}_Z(k)|} = \overline{|S_Z(k)|} - H(k) \overline{|S_S(k)|} \quad (2)$$

$E(k)$ corresponds to the difference between the envelope of the magnitude spectra of the optionally non-linearly distorted optionally smoothed target signal and its estimation. The optimization problem is observed separately for each frequency component k . The simplest realisation of the spectral filter $H(k)$ would be determined by the two envelopes, with

$$H(k) = \frac{\overline{|S_z(k)|}}{\overline{|S_s(k)|}}. \quad (3)$$

Alternatively, a constraint of $H(k)$ is suggested through the introduction of a regularization parameter. The underlying intention is to prevent the filter amplification from rising disproportionately if the signal power of $\overline{|S_s|}$ is too weak and hence, background noise becomes audible or the system becomes perceptibly instable. If, for example, the spectral peaks of one time-block of $\overline{|S_z|}$ and $\overline{|S_s|}$ are not located in exactly the same frequency band, $H(k)$ will rise excessively in these bands in which $\overline{|S_z|}$ has a maximum and $\overline{|S_s|}$ has a minimum. To avoid this problem, a constraint for $H(k)$ is established through the frequency-dependent regularisation parameter $\beta(k)$, yielding

$$H(k) = \frac{\overline{|S_z(k)|} \overline{|S_s(k)|}}{\overline{|S_s(k)|}^2 + \beta(k)}. \quad (4)$$

Through positive real-valued $\beta(k)$, the filter amplification will not increase immoderately, even with a small value for $\overline{|S_s|}$, and hence, will prevent undesired signal peaks. The optimal values for $\beta(k)$ depend on the signal statistics to be expected, whereas a computation based on an estimation of the background noise power per frequency band is proposed inventively. The background noise power $P_g(k)$ can be estimated incorporating the time-averaged minimum statistics. The regularisation parameter $\beta(k)$ is proportional to the rms value of

the background noise power, according to: $\beta(k) = c \cdot [P_g(k)]^{\frac{1}{2}}$, and c typically between 1 and 5.

An alternative implementation of H is proposed specifically for quasi-stationary input signals. The envelopes of the magnitude spectra are first estimated without time-averaging and optionally non-linear distortion. Both modifications are considered during the determination of the filter coefficients, according to:

$$\overline{H(m,k)} = \left\{ \alpha \left[\frac{|S_z(m,k)| |S_s(m,k)|}{|S_s(m,k)|^2 + \beta(k)} \right]^\gamma + (1-\alpha) \overline{H(m-1,k)}^\gamma \right\}^{\frac{1}{\gamma}} \quad (5)$$

In equation (5), both the block index m and the frequency index k are indicated, since the computation simultaneously depends on both indices in this case. The parameters α and γ determine the behavior of the time-averaging or the non-linear distortion.

Calculations in the event of dropouts in a target signal

The possibilities for detecting a dropout are numerous and known very well in the prior art. For example, a status bit can be transmitted at a reserved position within the respective audio stream (e.g. between audio data frames), and continuously registered at the receiver side. It would also be conceivable to perform an energy analysis of the individual frames and to identify a dropout if it falls below a certain threshold. A dropout could also be detected through synchronization between transmitter and receiver.

If a dropout is detected in the target signal (e.g. as represented in Fig. 2 by a status bit "dropout y/n"; the dotted line denotes the status bit that is actually transmitted contiguously with the audio signal), the replacement signal must be generated using the lastly estimated filter coefficients and the substitution channel(s), and is directly fed to the output of the

concealment unit. During a dropout, the estimation of the filter coefficients is deactivated. Basically, the transition between target and replacement signal can be implemented by a switch, assuming any switching artefacts remaining inaudible. According to the invention, a cross-fade between the signals is proposed as being advantageous, but this requires a buffering of the target signal, hence inducing additional latency. In particularly delay-critical real-time systems that do not allow for any additional buffering, a cross-fade is not readily possible. In this case, an extrapolation of the target signal is proposed, for example by means of linear prediction. The cross-fade is carried out between the extrapolated target signal and the replacement signal by using the method according to the invention.

The replacement signal is finally generated through filtering of the substitution signal with the filter coefficients retransformed into the time domain. The inverse transformation of the filter coefficients $T^{-1}\{H\}$ should be carried out with the same method as the first transformation. Prior to the filtering, the filter impulse response is optionally time-limited by a windowing function $w(n)$ (e.g. rectangular, Hanning).

$$h_w(n) = w(n)T^{-1}\{H(k)\} \text{ or } \overline{h_w(n)} = w(n)T^{-1}\{\overline{H(k)}\}. \quad (6)$$

The impulse response $h_w(n)$ or $\overline{h_w(n)}$, respectively, must only be calculated once at the beginning of the dropout, since the continuous estimation of the filter coefficients is deactivated during the dropout. For the sample-wise determination of the replacement signal \hat{x}_Z , an appropriate vector of the substitution signal \mathbf{x}_S is necessary,

$$\hat{x}_Z(n) = h_w^T \mathbf{x}_S(n) \text{ or } \hat{x}_Z(n) = \overline{h_w}^T \mathbf{x}_S(n). \quad (7)$$

In some applications, the filtering can be performed in the frequency domain. Thus, the coefficients optionally windowed in the time domain are transformed back into the frequency domain, so that the replacement signal of a block is computed by:

$$\hat{\mathbf{x}}_Z(n) = T^{-1}\{H_w^*(k)X_S(k)\}. \quad (8)$$

Successive blocks are combined using methods such as overlap and add or overlap and save.

The replacement signal is continued beyond the end of the dropout to enable a cross-fade into the re-existing target signal.

Estimation of the time delay between target and substitution signal

In a particularly preferred embodiment of the present concealment method, the time-alignment of target and replacement signal can be improved, too. Therefore, a time delay is estimated, parallel to the spectral filter coefficients, that takes two components into account. On the one hand, the delay of the replacement signal resulting from the filtering process must be compensated for, $\tau_1 = \frac{L_{Filter}}{2}$. On the other hand, a time delay τ_2 between target and substitution channel originates due to the spatial arrangement of the respective microphones. This can be estimated, for example, by means of the generalized cross-correlation (GCC) that requires the computation of complex short-term spectra. In a preferred implementation, the short-term DFT employed for the estimation of the concealment filter can be exploited, too, obviating additional computational complexity. (For more information about the characteristics of the GCC, see especially Carter, G. C.: "Coherence and Time Delay Estimation"; Proc. IEEE, Vol. 75, No. 2, Feb. 1987; and Omologo M., Svaizer P.: "Use of the Crosspower-Spectrum Phase in Acoustic Event Location"; IEEE Trans. on Speech and Audio Processing, Vol. 5, No. 3, May 1997. The entire disclosures are incorporated into this specification by virtue of reference.) The GCC is calculated using inverse Fourier transform of the estimated generalized cross-power spectral density (GXPSD), which is defined by:

$$\Phi_{G,ZS}(k) = G(k) X_Z(k) X_S^*(k) \quad (9)$$

(again, in equations 9-12, the block index m is omitted.)

In equation (9), $X_Z(k)$ and $X_S(k)$ are the DFTs of a block of the target or substitution channel, respectively; * denotes complex conjugation. $G(k)$ represents a pre-filter the aim of which is explained in the following.

The time delay τ_2 is determined by indexing the maximum of the cross-correlation. The detection of the maximum can be improved by approximating its shape to a delta function.

The pre-filter $G(k)$ directly affects the shape of the GCC and thus, enhances the estimation of τ_2 . A proper realisation denotes the phase transform filter (PHAT):

$$G_{PHAT}(k) = \frac{1}{|X_Z(k)X_S^*(k)|}. \quad (10)$$

This results in the GXPSD with PHAT filter:

$$\Phi_{G,ZS}(k) = \frac{X_Z(k)X_S^*(k)}{|X_Z(k)X_S^*(k)|} = \frac{\Phi_{ZS}(k)}{|X_Z(k)X_S^*(k)|}, \quad (11)$$

where Φ_{ZS} : cross-power spectral density of target and substitution signal.

Another possibility is offered by the complex coherence function whose pre-filter can be calculated from the power density spectra, yielding:

$$\Gamma_{ZS}(k) = \frac{\Phi_{ZS}(k)}{\sqrt{\Phi_{ZZ}(k)\Phi_{SS}(k)}} \quad (12)$$

Φ_{ZZ} : auto-power spectral density of the target signal,

Φ_{SS} : auto-power spectral density of the substitution signal.

The transformation of the signals into the frequency domain is usually implemented by means of short-term DFT. The block length must, on the one hand, be selected large enough in order to facilitate peaks in the GCC that are detectable for the expected time delays but, on the other hand, excessive block lengths lead to increased need for storage capacity. To adequately track variations of the time delay τ_2 , time-averaging of the GXPSD or of the complex coherence function is proposed (e.g. by exponential smoothing).

$$\overline{\Phi_{G,ZS}(m,k)} = \mu \frac{\Phi_{ZS}(m,k)}{|X_Z(m,k)X_S^*(m,k)|} + (1-\mu)\overline{\Phi_{G,ZS}(m-1,k)}, \quad (13)$$

$$\overline{\Gamma_{ZS}(m,k)} = \nu \frac{\Phi_{ZS}(m,k)}{\sqrt{\Phi_{ZZ}(m,k)\Phi_{SS}(m,k)}} + (1-\nu)\overline{\Gamma_{ZS}(m-1,k)}. \quad (14)$$

In equations (13) and (14), m refers to the block index. The smoothing constants are designated with μ and ν . These must be adapted to the jump distance of the short-term

DFT and the stationarity of τ_2 in order to obtain the best possible estimation of the coherence function or the generalized cross-power spectral density, respectively.

After the retransformation into the time domain and the detection of the maximum of the GCC, the entire time delay element between target and replacement signal can be formulated by

$$\Delta\tau = \tau_2 - \tau_1. \quad (15)$$

The individual processing steps are summarized in a block diagram in Fig. 2 for one target and one substitution signal. The transition between target and replacement signal or vice-versa is depicted as a simple switch in the graphic; as has already been mentioned, a cross-fade of the signals is recommendable.

The inventive notion of a multi-channel setup with more than two channels is depicted Fig. 3. Depending on which channel is affected by dropouts and hence becomes the target channel, the substitution signal is generated with the remaining intact channels. The discrete blocks of Fig. 3 correspond to the following processing steps:

- 21 Selection of the substitution channel(s)
- 22 Calculation of the filter coefficients
- 23 Application of a time delay
- 24 Generation of a replacement signal

In the uppermost row of Fig. 3, a replacement signal is generated for channel 1, which is afflicted by dropouts. To achieve this, either one, several, or all of the channels 2 to 7 can be used. The second row corresponds to the reconstruction of channel 2, etc.

Fig. 4 shows a schematic of the basic algorithm in combination with the expansion stage (i.e. time delay estimation) to illustrate the mutual dependencies of the individual processing

steps. To simplify the block diagram, parallel signals (DFT blocks) or (spectral) mappings derived thereof are merged into one (solid) line, the number of which is indicated by K or $K - 1$, respectively. The dotted connections denote the transfer or input of parameters. The first selection of the substitution channels is done in the block labeled "selector" according to the GXPSD. On the one hand, this affects the computation of the envelopes of the magnitude spectra of the substitution signal and, on the other hand, it is needed for the weighted superposition of the same. The second selection criterion is offered by the time delay τ_2 . The status bits of the channels are not depicted explicitly, but their verification is considered in relevant signal-processing blocks. Additionally, the particular determination of the target signal can be omitted from this illustration.

Hardware implementation

According to the invention, the method for dropout concealment works as an independent module and is intended for installation into a digital signal processing chain, wherein the software-specified algorithm is implemented on a commercially available digital signal processor (DSP), preferably a special DSP for audio applications. Accordingly, for each channel of a multi-channel arrangement, an appropriate device, such as exemplarily depicted in Fig. 5, is necessary that preferably may be integrated directly into the apparatus for receiving and decoding the transmitted digital audio data.

The apparatus for dropout concealment is equipped with a primary audio input that adopts the digital signal frames from the receiver unit and temporarily stores them in a storage unit 25. The apparatus is equipped with at least one secondary audio input, optionally several secondary audio inputs, at which the digital data of the substitution channel(s) are available and likewise stored temporarily in one, optionally several, storage unit(s) 25.

In addition, the device features an interface for the transmission of control data such as the status bit of the signal frames (dropout y/n) or an information bit for the selection of the substitution channel(s), the latter requiring (a) a bidirectional data line and (b) a temporary storage unit 25.

In order to forward the original or concealed data frames of the primary channel, the apparatus is equipped with an audio output. A separate storage unit for the data blocks to be output is not necessary, since they can be stored as needed in the storage unit of the input signal.

Patent claims:

1. Method for the concealment of dropouts in one or more channels (Z) of a multi-channel arrangement comprising at least two channels (Z, S), wherein in the event of a dropout in one channel (Z) a replacement signal is generated with the aid of at least one error-free channel (S), characterised in that, during the error-free signal transmission of the channels (Z, S) a mapping of the transmitted signals (x_z, x_s) into the frequency domain is performed, the magnitude spectra ($|S_z|, |S_s|$) being determined, spectral filter coefficients (H) are calculated that relate the magnitude spectrum ($|S_z|$) of a channel (Z) to the magnitude spectrum ($|S_s|$) of at least one other channel (S), and that in the event of the dropout of a channel (Z) the replacement signal is generated by application of filter coefficients (H), computed prior to the dropout, to a substitution signal which consists of at least one error-free channel (S).
2. Method according to Claim 1, characterised in that the magnitude spectra ($|S_z|, |S_s|$) are distorted non-linearly prior to the calculation of the filter coefficients (H).
3. Method according to one of Claims 1 or 2, characterised in that the magnitude spectra ($|S_z|, |S_s|$) are time-averaged prior to the calculation of the filter coefficients (H).
4. Method according to one of Claims 1 to 3, characterised in that the filter coefficients (H) are calculated by minimizing the difference between the optionally non-linearly distorted and/or time-averaged magnitude spectrum ($|S_z|$) of a channel (Z), and an optionally non-linearly distorted and/or time-averaged magnitude spectrum ($|S_s|$) of at least one other channel (S) filtered with the filter coefficients (H).

5. Method according to one of Claims 1 to 4, characterised in that the calculation of the filter coefficients (H) is done through the quotient of the magnitude spectra ($|S_z|$, $|S_s|$), according to:

$$H(k) = \frac{|S_z(k)|}{|S_s(k)|}.$$

6. Method according to one of Claims 1 to 5, characterised in that a regularisation of the filter coefficients (H) is carried out with the aid of a frequency-dependent parameter $\beta(k)$.

7. Method according to Claim 6, characterised in that the regularisation is accomplished,

according to the formulation
$$H(k) = \frac{|S_z(k)| |S_s(k)|}{|S_s(k)|^2 + \beta(k)}.$$

8. Method according to Claim 7, characterised in that the estimation of $\beta(k)$ is achieved via the rms value of the background noise level $P_g(k)$, where

$$\beta(k) = c \cdot [P_g(k)]^{\frac{1}{2}},$$
 the factor c facilitating an improved adaptation with preferred values of $c = 1 \dots 5$.

9. Method according to one of Claims 1 to 9, characterised in that the calculation of the envelopes of the magnitude spectra are obtained by means of the short-term DFT of short block length.

10. Method according to one of Claims 1 to 9, characterised in that the envelopes of the magnitude spectra can be calculated incorporating the magnitude spectra of a wavelet transformation, or the rms (per channel) of a gammatone filter bank, or linear prediction with subsequent sampling of the magnitude of the spectral envelopes of a signal frame (represented by the synthesis filter), or a real cepstral analysis with

subsequent retransformation of the cepstral domain into the frequency domain and taking the antilogarithm, or a short-term DFT with maximum detection and interpolation of the magnitude spectra, respectively.

11. Method according to Claim 3, characterised in that the time-averaging of a magnitude spectrum ($|S_Z|$, $|S_S|$) incorporates exponential smoothing with a smoothing constant (α).

12. Method according to Claim 3, characterised in that the time-averaging of a magnitude spectrum ($|S_Z|$, $|S_S|$) is implemented by means of a moving average filter.

13. Method according to Claims 2 and 3, characterised in that the non-linear distortion and the time-averaging of a magnitude spectrum ($|S_Z|$, $|S_S|$) obeys either formulation

$$\overline{|S_Z(m)|} = \left\{ \alpha |S_Z|^\gamma + (1-\alpha) \overline{|S_Z(m-1)|}^\gamma \right\}^{\frac{1}{\gamma}} \quad \text{or} \quad \overline{|S_S(m)|} = \left\{ \alpha |S_S|^\delta + (1-\alpha) \overline{|S_S(m-1)|}^\delta \right\}^{\frac{1}{\delta}}$$

where α refers to the smoothing constant in the range of $0 < \alpha \leq 1$, m refers to the block index and γ , δ refers to the distortion exponents for the magnitude spectra ($|S_Z|$, $|S_S|$).

14. Method according to Claim 2, characterised in that the non-linear distortion is achieved through the logarithmic and exponential function, where

$$\overline{|S_Z(m)|} = e^{\left\{ \alpha \ln\{|S_Z|\} + (1-\alpha) \ln\{\overline{|S_Z(m-1)|}\} \right\}} \quad \text{and} \quad \overline{|S_S(m)|} = e^{\left\{ \alpha \ln\{|S_S|\} + (1-\alpha) \ln\{\overline{|S_S(m-1)|}\} \right\}}.$$

15. Method according to one of Claims 1 to 4, characterised in that the calculation of the filter coefficients (H) is carried out by way of the time-averaging of the coefficients

instead of the time-averaging of the spectral envelopes, according to the formulation

$$\overline{H(m,k)} = \left\{ \alpha \left[\frac{|S_Z(m,k)| |S_S(m,k)|}{|S_S(m,k)|^2 + \beta(k)} \right]^\gamma + (1-\alpha) \overline{H(m,k)}^\gamma \right\}^{\frac{1}{\gamma}} .$$

16. Method according to one of Claims 1 to 15, characterised in that the filter coefficients (H) are transformed into time domain, and the filter impulse response is bounded in time domain applying a windowing function.
17. Method according to one of Claims 1 to 16, characterised in that the replacement signal is generated through the filtering of the error-free substitution channel in the time domain.
18. Method according to one of Claims 1 to 16, characterised in that the bounded filter impulse response is brought back into frequency domain, and the filtering of the substitution signal is performed in frequency domain.
19. Method according to one of Claims 1 to 18, characterised in that the transition between the target signal and the replacement signal takes place using a cross-fade.
20. Method according to Claim 19, characterised in that an extrapolation by means of a linear prediction filter is used for the implementation of the cross-fade without buffering and hence without additional signal delay.
21. Method according to one of Claims 1 to 20, characterised in that a time delay (τ_2) between the signals (x_Z, x_S) transmitted on the channels (Z, S) is determined from the magnitude spectra ($S_Z, S_S; X_Z, X_S$) of two channels, and is applied as a time delay to the replacement signal.

22. Method according to Claim 21, characterised in that the time delay (τ_2) is determined from the maximum of the generalized cross-correlation of the signals (x_z, x_s).

23. Method according to Claims 21 and 22, characterised in that the time delay (τ_2) is reduced by the time delay (τ_1) that is caused by the filtering of the substitution signal (x_s) with the time domain filter coefficients (h_w), yielding a new time delay $\Delta\tau = \tau_2 - \tau_1$ that this is applied to the replacement signal.

24. Method according to Claims 22 and 23, characterised in that the generalized cross-correlation is determined from the generalized cross-power spectral density $\Phi_{G,Zs}(k) = G(k)X_z(k)X_s^*(k)$ through inverse transformation of the latter into the time domain; ($G(k)$) refers to a pre-filter and (X_z, X_s) refers to the complex spectra of the signals (x_z, x_s).

25. Method according to Claim 24, characterised in that the pre-filter ($G(k)$) is the phase

transform filter
$$G_{PHAT}(k) = \frac{1}{|X_z(k)X_s^*(k)|}.$$

26. Method according to Claims 22 and 23, characterised in that the generalized cross-correlation is determined by inverse transformation of the coherence function

$$\Gamma_{Zs}(k) = \frac{\Phi_{Zs}(k)}{\sqrt{\Phi_{Zz}(k)\Phi_{Ss}(k)}}$$
 into the time domain, where $\Phi_{Zs}(k) = X_z(k)X_s^*(k)$ and

$\Phi_{Zz}(k)$ and $\Phi_{Ss}(k)$ refer to the auto-power spectral densities of the two signals (Z, S).

27. Method according to one of Claims 21 to 26, characterised in that the frequency spectra (X_z, X_s) of the signals (x_z, x_s) are determined by means of short-term DFT.

28. Method according to one of Claims 21 to 27, characterised in that, prior to the transformation into the time domain, the generalized cross-power spectral density or the coherence function is preferably time-averaged through exponential smoothing.

29. Method according to one of Claims 1 to 28, characterised in that, a signal $x_j(n)$ is selected as a substitution signal, whose frequency-averaged version of the coherence function, $\chi(j) = \frac{1}{N} \sum_{k=0}^{N-1} \left| \overline{\Gamma_{zS,j}(k)} \right|$, is maximum, according to: $x_s(n) = x_j(n)$ with $J = \arg \max_j \chi(j)$.

30. Method according to Claims 1 to 28, characterised in that the substitution signal is composed of several weighted signals.

31. Method according to Claim 30, characterised in that a superposition of several channels to form one substitution channel is implemented, according to the

formulation
$$x_s(n) = \frac{\sum_{j \in \tilde{J}} \left\{ \chi(j) \cdot x_j(n - \Delta\tau_j) \right\}}{\sum_{j \in \tilde{J}} \chi(j)}$$
, where \tilde{J} represents the set of the

indices of the potential channels. The superposition considers all time delays ($\Delta\tau_j$), too.

32. Method according to Claim 31, characterised in that the size of \tilde{J} can be delimited by the user.

33. Method according to Claims 31 and 32, characterised in that the size of \tilde{J} is restricted to channels whose frequency-averaged values of the coherence function (with the target channel) $\chi(j)$ exceed a threshold value Θ , according to:

$$\tilde{J} = \left\{ j \mid (1 \leq j \leq K-1) \wedge [\chi(j) > \Theta] \right\}.$$

34. Method according to Claims 31 and 32, characterised in that the size of \tilde{J} is restricted to a maximum number of M channels, according to:

$$\tilde{J} = \left\{ j_i \mid (1 \leq j_i \leq K-1) \wedge (1 \leq i \leq M) \wedge \left[\chi(j_i) > \chi(l), \forall l \in \{1, \dots, K-1\} \setminus \{j_1, \dots, j_M\} \right] \right\}.$$

35. Method according to Claims 31 to 34, characterised in that the criteria threshold value Θ and maximum number M are jointly taken into account, according to:

$$\tilde{J} = \left\{ j_i \mid (1 \leq j_i \leq K-1) \wedge (1 \leq i \leq M) \wedge (\chi(j_i) > \Theta) \wedge \left[\chi(j_i) > \chi(l), \forall l \in \{1, \dots, K-1\} \setminus \{j_1, \dots, j_M\} \right] \right\}$$

36. Method according to one of Claims 1 to 28, characterised in that different substitution signals are used for different frequency bands of the replacement signal.

37. Method according to Claim 36, characterised in that, for each frequency band k , an appropriate band-pass-filtered version of that signal $x_{J,k}(n)$ is selected as a substitution signal whose value of the (time-averaged) coherence function $\left| \overline{\Gamma_{zS,j}(k)} \right|$ with the signal to be replaced has a maximum value in the respective frequency band k prior to the dropout, according to: $x_{S,k}(n) = x_{J,k}(n)$, where $J = \arg \max_j \left| \overline{\Gamma_{zS,j}(k)} \right|$.

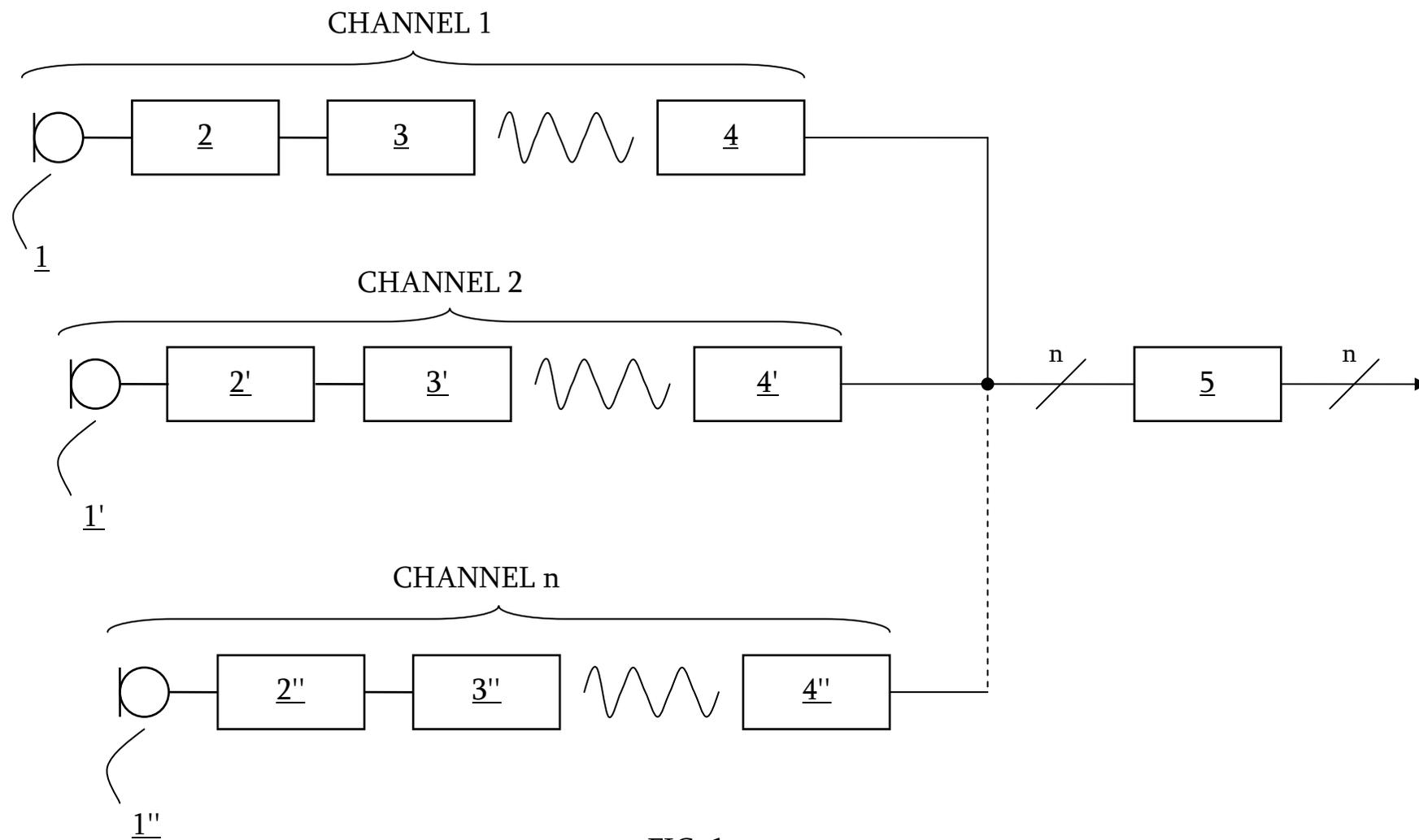


FIG. 1

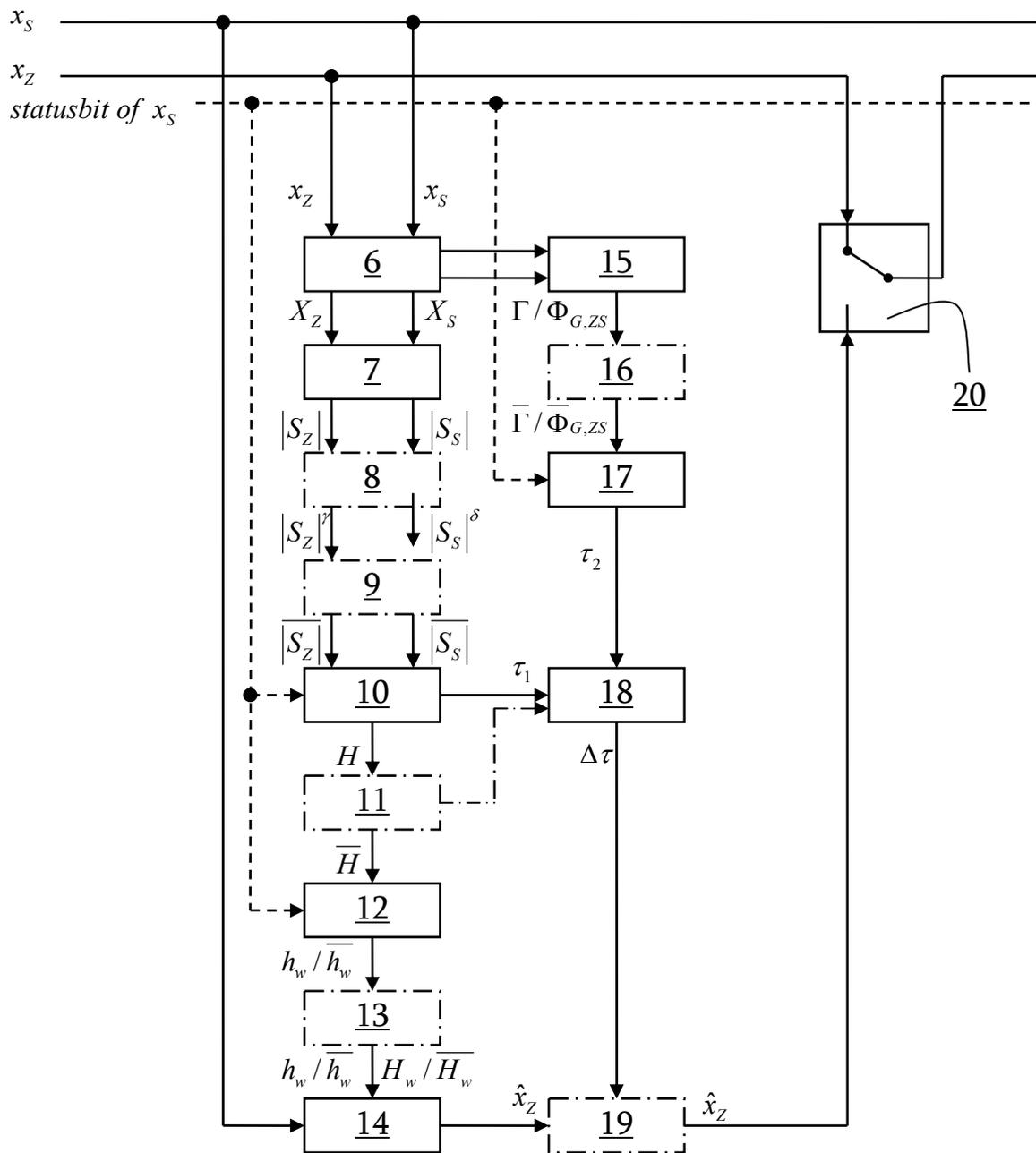


FIG. 2

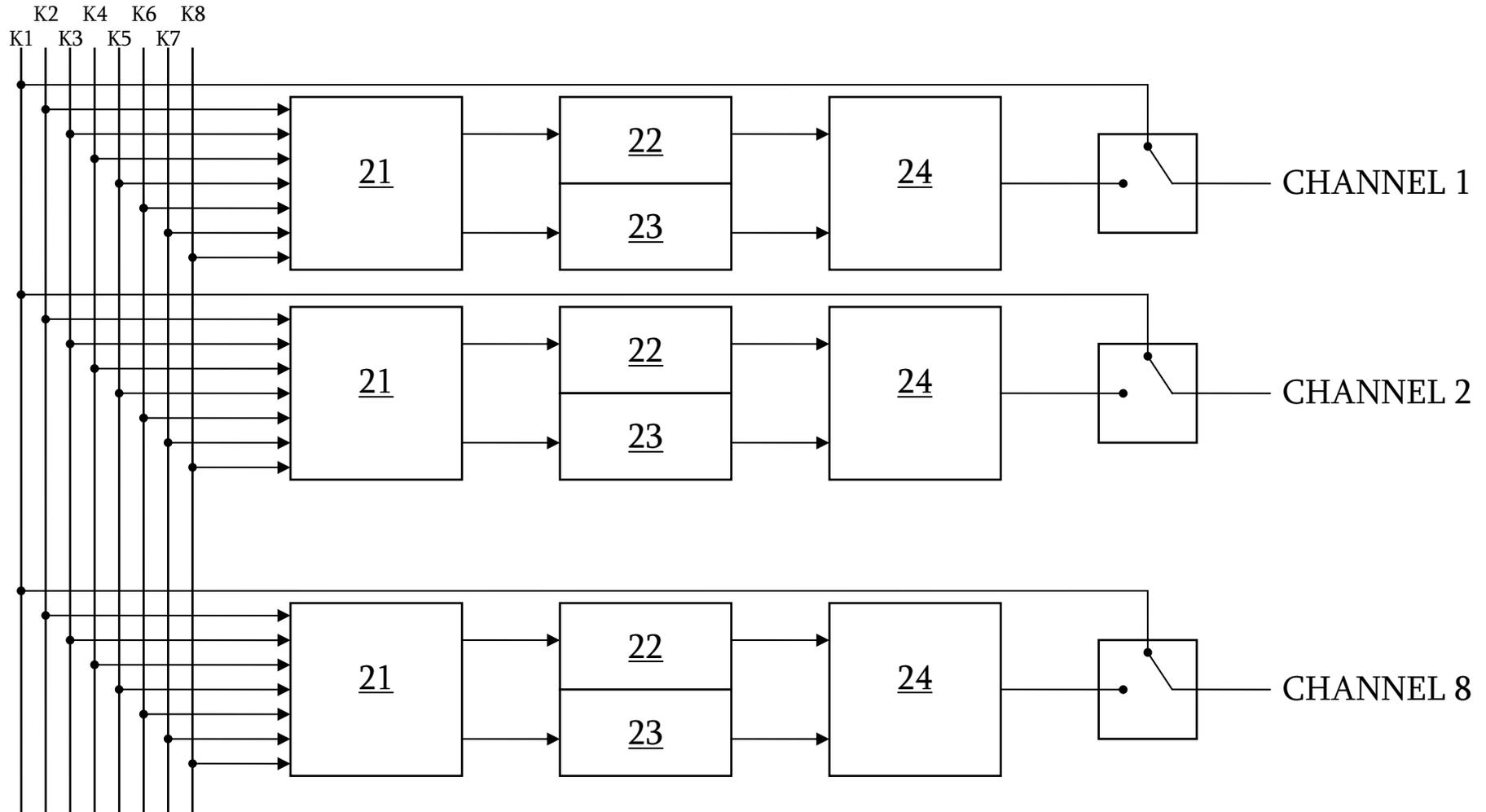


FIG. 3

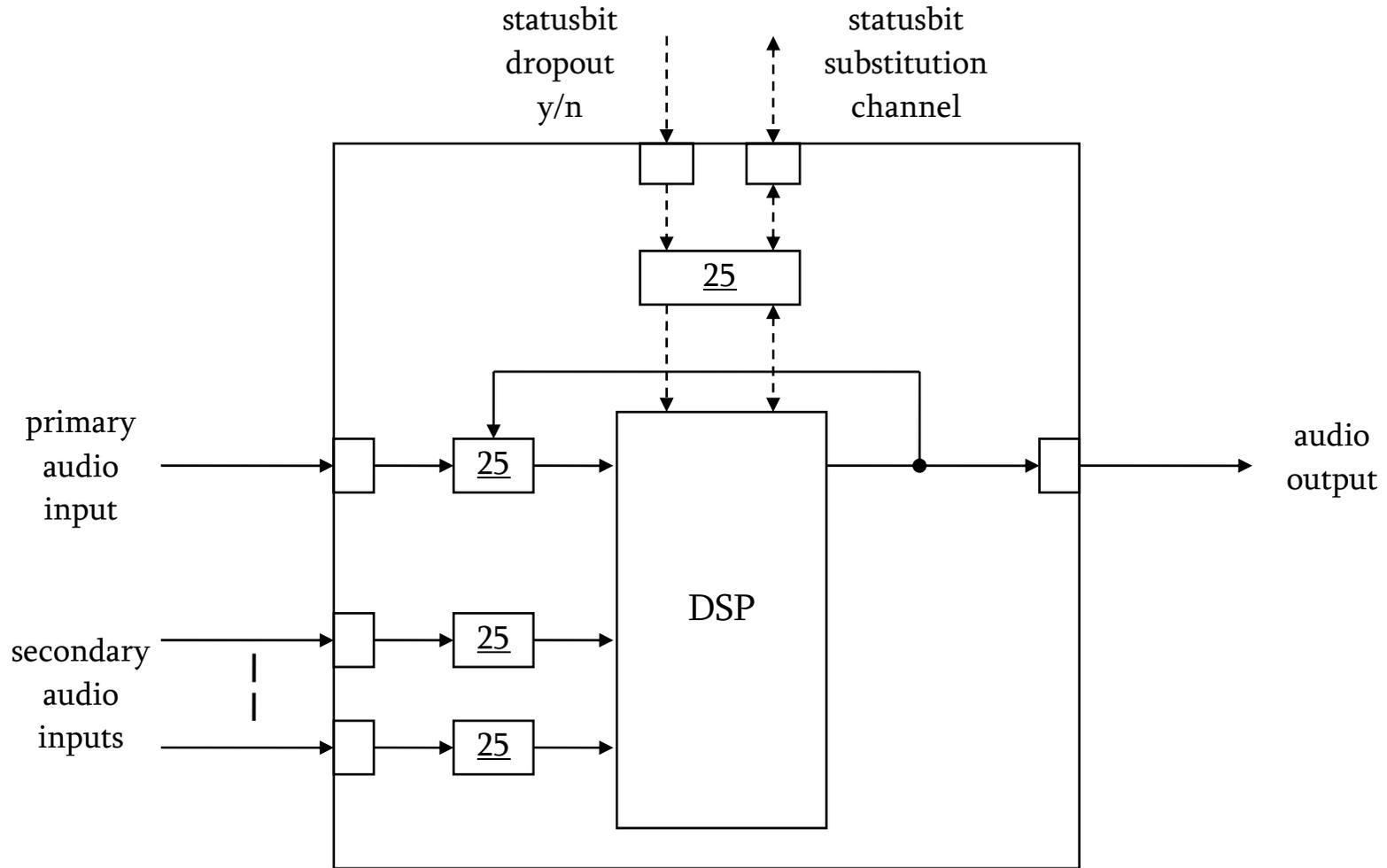


FIG. 5

Anhang C

Abkürzungsverzeichnis

ADC	Analog-Digital-Konverter
AKF	Autokorrelationsfolge
AMDF	<i>average magnitude difference function</i>
AP	affine Projektion
AR	autoregressiv (s. Polstellen-Filter)
ARMA	autoregressiv, <i>moving average</i> (s. Pol-Nullstellen-Filter)
CMNDF	<i>cumulative mean normalized difference function</i>
DF	Differenzfunktion
DFT	diskrete Fouriertransformation
DROCON	<i>dropout concealment</i>
E-DROCON	Verfahren zur Ausfallsverschleierung Einkanalssystemen
GCC	<i>generalised cross correlation</i>
GXPSD	<i>generalised cross power spectral density</i>
IP	<i>internet protocol</i>
LMS	<i>least mean square</i>
LP	lineare Prädiktion
LRPFF	lineares Rückwärtsprädiktionsfehlerfilter

LSF	<i>line spectral frequencies</i>
LVPPF	lineares Vorwärtsprädiktionsfehlerfilter
MA	<i>moving average</i> (s. Nullstellen-Filter)
M-DROCON	Verfahren zur Ausfallsverschleierung in Mehrkanalsystemen
MIMO	<i>multiple input multiple output</i>
MSD	<i>minimum structure distance</i>
NLMS	<i>normalised least mean square</i>
PDA	<i>pitch determination algorithm</i>
PHAT	<i>phase transform</i>
PS	Periodizitätsschätzung
SV	Signalverarbeitung
TCP	<i>transmission control protocol</i>
UDP	<i>user data protocol</i>
VP	Versuchsperson

Anhang D

Literaturverzeichnis

- [Areal 2000] Areal R., Kovacic J., and Goyal V. K.: "Multiple description perceptual audio coding with correlating transforms"; IEEE Trans. on Speech and Audio Proc., Vol. 8., No. 2, March 2000.
- [Bäckström 2004] Bäckström T.: "Linear Predictive Modelling of Speech – Constraints and Line Spectrum Pair Decomposition"; Dissertation, Helsinki University of Technology, 2004.
- [Barros 2002] Barros J., Oikonomidis I.: "Wireless transmission of packet audio using multiple description"; IEEE PIMRC Symposium'02, Vol 5., pp. 2452-2455, Sept. 15-18, 2002.
- [Bartels 2006] Bartels R. H., Beatty J. C., and Barsky B. A.: "An Introduction to Splines for Use in Computer Graphics and Geometric Modeling"; Morgan Kaufmann Publishers Inc., San Mateo, California, ISBN 1-558-60400-6, 2006.
- [Bortz 2005] Bortz J.: "Statistik für Human- und Sozialwissenschaftler"; 6. Auflage, Springer Medizin Verlag, Heidelberg, ISBN 3-540-21271-X, 2005.
- [Brandstein 2001] Brandstein M., and Ward D.: "Microphone Arrays"; Springer Verlag, Berlin Heidelberg New York, ISBN 3-540-41953, 2001.
- [Bruhn 2005a] Bruhn S.: "Channel Signal Concealment In Multi-channel Audio Systems"; US Patent Application Publication, US Publication Nr: US 2005/0182996 A1, US Publication Date: 18.06.2005.
- [Bruhn 2005b] Bruhn S.: "Channel Signal Concealment In Multi-channel Audio Systems"; World Intellectual Property Organization, Int. Publication Nr: WO 2005/059898 A1, Int. Publication Date: 30.06.2005.

- [Bruhn 2006] Bruhn S.: "Channel Signal Concealment In Multi-channel Audio Systems"; European Patent Office, EU Publication Nr: EP 1 649 452 A1, EU Publication Date: 26.04.2006.
- [Carter 1987] Carter G. C.: "Coherence and Time Delay Estimation"; Proc. IEEE, Vol. 75, No. 2, Feb. 1987.
- [Daubechies 1992] Daubechies I.: "Ten Lectures on Wavelets"; Society for Industrial and Applied Mathematics, Capital City Press, Vermont, ISBN 0-89871-274-2, 1992.
- [de Cheveigne 2002] de Cheveigne A., and Kawahara H.: "YIN, a fundamental frequency estimator for speech and music"; J. Acoust. Soc. Am., Vol. 111(4), pp. 1917-1930, April 2002.
- [Deller 2000] Deller J. R., Hansen J. H. L., and Proakis J. G.: "Discrete-Time Processing of Speech Signals"; IEEE Press, Piscataway, New Jersey, ISBN 0-7803-5386-2, 2000.
- [Esquef 2002] Esquef P. A. A., Karjalainen M., and Välimäki V.: "Detection of Clicks in Audio Signals using Warped Linear Prediction"; Proc. IEEE DSP'02, Vol. 2, pp. 1085-1088, July 1-3, 2002.
- [Esquef 2003] Esquef P. A. A., Välimäki V., Roth K., and Kauppinen I.: "Interpolation of Long Gaps in Audio Signals using the Warped Burg's Method"; Proc. DAFX'03, Sept. 8-11, 2003.
- [Esquef 2006] Esquef P. A. A., and Biscainho L. W. P.: "An Efficient Model-Based Multirate Method for Reconstruction of Audio Signals Across Long Gaps"; IEEE Trans. on Audio, Speech and Language Proc., Vol. 14, No. 4, July 2006.
- [Etter 1996] Etter W.: "Restauration of a Discrete-Time Signal Segment by Interpolation Based on the Left-Sided and Right-Sided Autoregressive Parameters"; IEEE Trans. on Signal Processing, Vol. 44, No. 5, May 1996.
- [Gay 1993] Gay S. L.: "A Fast Converging, Low Complexity Adaptive Filtering Algorithm"; IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, Vol. 1, pp. 04-07, Oct. 17-20, 1993.
- [Gay 1995] Gay S. L., and Tavathia S.: "The Fast Affine Projection Algorithm"; Proc. IEEE ICASSP'95, Vol. 5, pp. 3023-3026, May 09-12, 1995.
- [Godsill 1997] Godsill S. J.: "Robust Modelling of Noisy ARMA Signals"; Proc. IEEE ICASSP'97, Vol. 5, pp. 3797-3800, April 21-24, 1997.
- [Godsill 1998] Godsill S. J., and Rayner P. J. W.: "Digital Audio Restauration – a statistical model based approach"; Springer Verlag, ISBN 3-540-76222-1, 1998.
- [Goodman 1986] Goodman D. J., Lockhart G. B., Wasem O. J., and Wong W.: "Waveform Substitution Techniques for Recovering Missing Speech Segments in Packet Voice

-
- Communications"; IEEE Trans. on Acoustics, Speech and Signal Processing, Vol. ASSP-34, No. 6, Dec. 1986.
- [Härmä 2001a] Härmä A.: "Linear Predictive Coding With Modified Filter Structures"; IEEE Trans. on Speech and Audio Processing, Vol. 9, No. 8, Nov. 2001.
- [Härmä 2001b] Härmä A.: "Frequency-Warped Autoregressive Modelling and Filtering"; Dissertation HUT (Laboratory of Acoust. and Audio Sig. Proc.), Report 61, 2001.
- [Haykin 2002] Haykin S.: "Adaptive Filter Theory"; Prentice Hall Inc., Englewood Cliffs, New Jersey, ISBN 0-13-048434-2, 2002.
- [Herre 1995] Herre J.: "Fehlerverschleierung bei spektral codierten Audiosignalen"; Dissertation, Universität Erlangen, 1995.
- [Hess 1983] Hess W.: "Pitch Determination of Speech Signals"; Springer-Verlag Berlin, ISBN 3-540-11933-7, 1983.
- [Horak 1998] Horak C.: "Entwicklung eines Audiocodierverfahrens auf einem Signalprozessor"; Diplomarbeit, Institut f. Nachrichtentechnik und Hochfrequenztechnik, Technische Universität Wien, 1998.
- [Irino 2001] Irino T., and Patterson R. D.: "A compressive gammachirp auditory filter for both physiological and psychophysical data"; J. Acoust. Soc. Am., Vol. 109, pp. 2008-2022, 2001.
- [ITU-R 1997] Int. Telecommunication Union: "Methods For The Subjective Assessment Of Small Impairments In Audio Systems Including Multichannel Sound Systems"; ITU-R BS.1116-1, 1994-1997.
- [ITU-R 1998] Int. Telecommunication Union: "Methods of Objective Measurements Of Perceived Audio Quality"; ITU-R BS.1387-1, 1998-2001.
- [ITU-R 2003] Int. Telecommunication Union: "General Methods For Subjective Assessment Of Sound Quality"; ITU-R BS.1284-1, 1997-2003.
- [ITU-T 2000] Int. Telecommunication Union: "One-Way Transmission Time"; ITU-T G.114, 2000.
- [Janssen 1986] Janssen A. J. E. M., Veldhuis R. N. J., and Vries L. B.: "Adaptive Interpolation Of Discrete-Time Signals That Can Be Modeled As Autoregressive Processes"; IEEE Trans. on Acoustics, Speech and Signal Processing, Vol. ASSP-34, No. 2, April 1986.
- [Karadimou 2006] Karadimou K., Mouchtaris A., and Tsakalides P.: "Packet Loss Concealment for Multichannel Audio Using the Multiband Source/Filter Model"; Proc. 40th Annual Asilomar Conf. on Signals, Systems and Computers, Oct. 29 – Nov. 01, 2006.
-

- [Karjalainen 1997] Karjalainen M., Härmä A., Laine U. K., and Houpaniemi J.: "Warped Filters and their Audio Applications"; Proc. IEEE ASSP Workshop'97, Oct. 19-22, 1997.
- [Kauppinen 2001] Kauppinen I., Kauppinen J., and Saarinen P.: "A Method For Long Interpolation Of Audio Signals"; J. Aud. Eng. Soc., Vol. 49, No. 12, Dec. 2001.
- [Kreyszig 1993] Kreyszig E.: "Advanced Engineering Mathematics"; J. Wiley & Sons, New York, ISBN 0-471-59989-1, 1993.
- [Kruskal 1952] Kruskal W. H., and Wallis W. A.: "Use of ranks in one-criterion variance analysis"; J. American Statistics Association, Vol. 47, 1952.
- [Lagrange 2005] Lagrange M., and Marchand S.: "Long Interpolation of Audio Signals Using Linear Prediction in Sinusoidal Modeling"; J. Audio Eng. Soc., Vol. 53, No. 10, Oct. 2005.
- [Liao 2001] Liao W.-T., Chen J.-C., and Chen M.-S.: "Adaptive Recovery Techniques for Real-Time Audio Streams"; Proc. IEEE Int. Conf. of Computer and Communications Societies, INFOCOM'01, Vol. 2, April 22-26, 2001.
- [Lindblom 2002] Lindblom J., and Hedelin P.: "Packet Loss Concealment based on Sinusoidal Extrapolation"; Proc. IEEE ICASSP'02, Vol. 1, pp. 173-176, May 13-17, 2002.
- [Maher 1994] Maher R. C.: "A Method of Extrapolation of Missing Digital Audio Data"; J. Audio Eng. Soc., Vol. 42, No. 5, May 1994.
- [Makhoul 1975] Makhoul J.: "Linear Prediction – A Tutorial Review"; Proc. IEEE, Vol. 63, No. 4, April 1975.
- [Mann 1947] Mann H. B., and Whitney D. R.: "On a Test whether one of two Random Variables is Stochastically Larger than the other"; The Annals Of Mathematical Statistics, Vol. 18, 1947.
- [Marple 1987] Marple S. L. Jr.: "Digital Spectral Analysis With Applications"; Prentice Hall Inc., Englewood Cliffs, New Jersey, ISBN 0-13-214149-3, 1987.
- [Neumann 2001] Neumann O., and Schäffler M.: "Raummikrofonierung bei Mehrkanal-Musikaufnahmen"; Diplomarbeit, FH Stuttgart, 2001.
- [Niedzwiecki 1993] Niedzwiecki M.: "Statistical Reconstruction of Multivariate Time Series"; IEEE Trans. on Signal Proc., Vol. 41, No. 1, Jan. 1993.
- [Niedzwiecki 2001] Niedzwiecki M., and Cisowsky K.: "Smart Copying – A New Approach to Reconstruction of Audio Signals"; IEEE Trans. on Signal Processing, Vol. 49, No. 10, Oct. 2001.

- [Ofir 2005] Ofir H., and Malah D.: "Packet Loss Concealment for Audio Streaming Based on the GAPES Algorithm"; Preprint 118th AES Convention, preprint 6334, May 28-31, 2005.
- [Omologo 1997] Omologo M., and Svaizer P.: "Use of the Crosspower-Spectrum Phase in Acoustic Event Location"; IEEE Trans. on Speech and Audio Processing, Vol. 5, No. 3, May 1997.
- [Oppenheim 1975] Oppenheim A. V., and Schaffer R. W.: "Digital Signal Processing"; Prentice Hall Inc., Englewood Cliffs, New Jersey, ISBN 0-13-214635-5, 1975.
- [Papaodysseus] Papaodysseus C. N., Koukoutsis E. B., and Triantafyllou C. N.: "Error Sources and Error Propagation in the Levinson-Durbin-Algorithm"; IEEE Trans. on Signal Processing, Vol. 41, No. 4, April 1993.
- [Parikh 2000] Parikh V. N., Chen J.-H., and Aguilar G.: "Frame Erasure Concealment Using Sinusoidal Analysis/Synthesis and Its Application to MDCT-Based Codecs";
- [Perkins 1998] Perkins C., Hodson O., and Hardman V.: "A Survey of Packet Loss Recovery Techniques for Streaming Audio"; IEEE Network, Vol. 12, Iss. 5, Sept./Oct. 1998.
- [Rahkila 1997] Rahkila M., Houpaniemi J.: "Real-Time Internet Audio – Problems and Solutions"; Preprint 102nd AES Convention, preprint 4477, March 22-25, 1997.
- [Rayner 1991] Rayner P. J. W., and Godsill S. J.: "The Detection and Correction of Artefacts in Degraded Gramophone Recordings";
- [Rombouts 2000] Rombouts G., and Moonen M.: "A fast exact frequency domain implementation of the exponentially windowed affine projection algorithm"; IEEE Adaptive Systems for Signal Processing, Communications and Control Symposium, Vol. 1, pp. 342-346, Oct. 01-04, 2000.
- [Roth 2004] Roth K., and Kauppinen I.: "Exponential Weighting Method for Sample-by-Sample Update of Warped AR-Model"; Proc. DAFX'04, Oct. 5-7, 2004.
- [Rupp 1998] Rupp M.: "A Family of Adaptive Filter Algorithms with Decorrelating Properties"; IEEE Trans. on Signal Processing, Vol. 46, No. 3, March 1998.
- [Schildbach 2002] Schildbach W., Herre J., Sieler M., Brandenburg K.-H., and Gerhäuser H.: "Verfahren zum Verschleiern von Fehlern in einem Audiodatenstrom"; Deutsches Patent- und Markenamt, Patentschrift DE 197 35 675 C 2, 12. 12. 2002.
- [Sinha 2003] Sinha R., Papadopoulos C., and Kyriakakis C.: "Loss concealment for multichannel streaming audio"; ACM Network and Operating System Support for Digital Audio and Video (NOSSDAV), June 2003.

- [Synnes 1999] Synnes K., Parnes P., and Schefström D.: "Robust Audio Transport using mAudio"; Research Report, Lulea University of Technology, Dept. of Computer Science, 1999.
- [Thiede 2000] Thiede T., Treurniet W. C., Bitto R., Schmidmer C., Sporer T., Beerends J. G., Colomes C., Keyhl M., Stoll G., Brandenburg K., and Feiten B.: "PEAQ – The ITU Standard for Objective Measurement of Perceived Audio Quality"; J. Audio Eng. Soc., Vol. 48, No. 1/2/3, Jan/Feb/March 2000.
- [Van Veen 1988] Van Veen B. D., and Buckley K. M.: "Beamforming: A Versatile Approach to Spatial Filtering"; IEEE ASSP Magazine, pp. 4-24, April 1998.
- [Vaseghi 1996] Vaseghi S. V.: "Advanced Signal Processing and Digital Noise Reduction"; Teubner Verlagsges., Leipzig, ISBN 3-519-06451-0, 1996.
- [Virolainen 2004] Virolainen J., and Lakaniemi A.: "Enhanced Error Concealment for Spatial Audio"; US Patent Application Publication, US Publication Nr: US 2004/0039464 A1, US Publication Date: 26.02.2004.
- [Wah 1998] Wah B. W., and Lin D.: "Transformation-based Reconstruction for Audio Transmission over the Internet";
- [Wah 2000a] Wah B. W., Su X., and Lin D.: "A Survey of Error Concealment Schemes for Real-Time Audio and Video Transmission over the Internet"; Proc. IEEE Int. Symposium on Multimedia Software Engineering, Dec. 2000.
- [Wah 2000b] Wah B. W., Lin D., and Su X.: "Streaming Real-Time Audio and Video Data with Transformation-based Error Concealment and Reconstruction";
- [Wasem 1988] Wasem O. J., Goodman D. J., Dvorak C. A., and Page H. G.: "The Effect of Waveform Substitution on the Quality of PCM Packet Communication"; IEEE Trans. on Acoustics, Speech and Signal Processing, Vol. 36, No. 3, March 1988.
- [Weerackody 2001] Weerackody V., Schuller G., and Lou H.-L.: "Streaming of Multimedia with Reduced Start-Up Delay";
- [Wiese 2002] Wiese D., and Sedlmeyer R.: "Error Concealment in Digital Transmissions"; US Patent Application Publication, US Publication Nr: US 6.490551 B2, US Publication Date: 03.12.2002.
- [Xu 2000] Xu A., Woszczyk W., Settel Z., Pennycook B., Rowe R., Galanter P., Bary J., Martin G., Corey J., and Cooperstock J. R.: "Real-Time Streaming of Multichannel Audio Data over Internet"; J. Aud. Eng. Soc., Vol. 48, No. 7/8, July/Aug. 2000.
- [Zöfel 1992] Zöfel P.: "Statistik in der Praxis"; 3. Auflage, UTB für Wissenschaft, Gustav Fischer Verlag, Stuttgart, ISBN 3-437-40289-7, 1992.

Ich danke allen, die mich bei der Abfassung dieser Arbeit unterstützt haben.

Im Besonderen gilt mein Dank Herrn Prof. Robert Höldrich für die Ermöglichung dieser Arbeit. Sein Interesse und kritisches Begutachten führten zu unzähligen aufschlussreichen und spannenden Fachdiskussionen, die sehr zur Erweiterung meines Fachwissens beigetragen haben.

Herrn Martin Opitz und Herrn Peter Schlager von AKG Acoustics danke ich speziell für die angenehme Atmosphäre während der Zusammenarbeit am Forschungsprojekt DROCON.

Ganz herzlich bedanken möchte ich mich bei Markus für seine fruchtbaren Denkanstöße, vor allem aber auch für seine freundschaftlichen Ratschläge.

Meiner Familie möchte ich schließlich für ihre Unterstützung von ganzem Herzen danken.